



CERTIFIED COPY OF  
PRIORITY DOCUMENT

## Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

**Aktenzeichen:** 199 32 230.9

**Anmeldetag:** 9. Juli 1999

**Anmelder/Inhaber:** BASF AG, Ludwigshafen/DE

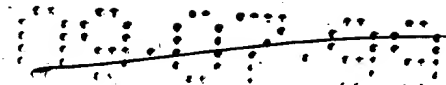
**Bezeichnung:** Corynebacterium Glutamicum Genes encoding  
Proteins involved in Membrane Synthesis and  
Membrane Transport

**IPC:** C 07 K 14/34

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 4. Juli 2002  
Deutsches Patent- und Markenamt  
Der Präsident  
Im Auftrag

BEST AVAILABLE COPY



## **CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT**

### **Abstract of the Disclosure**

5

Isolated nucleic acid molecules, designated MCT nucleic acid molecules, which encode novel MCT proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MCT nucleic acid molecules, and host cells into which the expression

10 vectors have been introduced. The invention still further provides isolated MCT proteins, mutated MCT proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of MCT genes in this organism.

## **CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT**

### **Background of the Invention**

5 Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic  
10 compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have  
15 been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

### **Summary of the Invention**

20 This invention provides novel nucleic acid molecules which may be used to identify or classify *Corynebacterium glutamicum* or related species of bacteria. *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The  
25 nucleic acid molecules therefore can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. While *C. glutamicum* itself is nonpathogenic, it is related to other *Corynebacterium* species, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria), which are important human pathogens. The ability to identify the presence of *Corynebacterium* species  
30 therefore also can have significant clinical relevance, e.g., diagnostic applications. Further, these nucleic acid molecules may serve as reference points for the mapping of the *C. glutamicum* genome, or of genomes of related organisms.

These novel nucleic acid molecules encode proteins, referred to herein as membrane construction and membrane transport (MCT) proteins. These MCT proteins  
35 are capable of, for example, performing a function involved in the metabolism (e.g., the biosynthesis or degradation) of compounds necessary for membrane biosynthesis, or of assisting in the transmembrane transport of one or more compounds either into or out of

the cell. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985);  
5 Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984), and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of  
10 the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell  
15 may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursors,  
20 cofactors, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may  
25 be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, MCT proteins of the  
30 invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which  
35 would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of



transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced.

5 This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of  
10 fine chemicals from *C. glutamicum* in large-scale fermentative culture.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as MCT proteins, which are capable of, for example, participating in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Nucleic acid  
15 molecules encoding an MCT protein are referred to herein as MCT nucleic acid molecules. In a preferred embodiment, the MCT protein participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of such proteins include those encoded by the genes set forth in Table 1.

20 Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs) comprising a nucleotide sequence encoding an MCT protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MCT-encoding nucleic acid (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid  
25 molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more  
30 preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein.

35 In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g.,

00.07.00

sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an MCT activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (e.g., an MCT fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MCT protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MCT protein by culturing the host cell in a suitable medium. The MCT protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an MCT gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated MCT sequence as a transgene. In another embodiment, an endogenous MCT gene within the genome of

the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered MCT gene. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

Still another aspect of the invention pertains to an isolated MCT protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated MCT protein or portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In another preferred embodiment, the isolated MCT protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes.

The invention also provides an isolated preparation of an MCT protein. In preferred embodiments, the MCT protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated MCT protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated MCT protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of MCT proteins also have one or more of the MCT bioactivities described herein.

00-07-99

The MCT polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MCT polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MCT protein alone. In other preferred embodiments, this fusion protein participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an MCT nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MCT nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MCT protein activity or MCT nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* metabolic pathways for cell membrane components or is modulated for the transport of compounds across such membranes, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MCT protein activity can be an agent which stimulates MCT protein activity or MCT nucleic acid expression. Examples of agents which stimulate MCT protein activity or MCT nucleic acid expression include small molecules, active MCT proteins, and nucleic acids encoding MCT proteins that have been introduced into the cell. Examples of agents which inhibit MCT activity or expression include small molecules and antisense MCT nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MCT gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be

modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

5

### Detailed Description of the Invention

The present invention provides MCT nucleic acid and protein molecules which are involved in the metabolism of cellular membrane components in *C. glutamicum* or in the transport of compounds across such membranes. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or optimization of a fatty acid biosynthesis protein has a direct impact on the yield, production, and/or efficiency of production of the fatty acid from modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the metabolism of cell membrane components results in alterations in the yield, production, and/or efficiency of production or the composition of the cell membrane, which in turn may impact the production of one or more fine chemicals). Aspects of the invention are further explicated below.

15  
20

#### 1. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research - Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, and all

25  
30  
35

other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

5    A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the  
10    nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L- amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids  
15    have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3<sup>rd</sup> edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosynthesis, are readily converted by simple biosynthetic pathways  
20    to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are  
25    interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout  
30    the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L- methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L- methionine are common feed additives. (Leuchtenberger, W. (1996) Amino acids -  
35    technical production and use, p. 466-502 in Rehm et al. (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-

acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E. (1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of  $\alpha$ -ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transfer of the side-chain  $\beta$ -carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. *Biochemistry* 3<sup>rd</sup> ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. *Biochemistry*, 3<sup>rd</sup> ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

*B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses*

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613. VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research - Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B<sub>1</sub>) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B<sub>2</sub>) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B<sub>6</sub>' (e.g., pyridoxine, pyridoxamine, pyridoxal-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced



either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of  $\beta$ -alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to  $\beta$ -alanine and for the condensation to panthothenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B<sub>5</sub>), pantetheine (and its derivatives) and coenzyme A.

10 Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the  $\alpha$ -ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

20 Corrinoids (such as the cobalamines and particularly vitamin B<sub>12</sub>) and porphyrines belong to a group of chemicals characterized by a tetrapyrrole ring system. The biosynthesis of vitamin B<sub>12</sub> is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

25 The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B<sub>6</sub>, pantothenate, and biotin. Only Vitamin B<sub>12</sub> is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

09.07.99

### C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) *Nucleotides and Related Compounds in Biotechnology* vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: *Progress in Nucleic Acid Research and Molecular Biology*, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*,

09.07.99

Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

#### D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in  $\alpha, \alpha$ -1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) *J. Japan* 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

## II. Membrane Biosynthesis and Transmembrane Transport

Cellular membranes serve a variety of functions in a cell. First and foremost, a membrane differentiates the contents of a cell from the surrounding environment, thus giving integrity to the cell. Membranes may also serve as barriers to the influx of hazardous or unwanted compounds, and also to the efflux of desired compounds. Cellular membranes are by nature impervious to the unfacilitated diffusion of hydrophilic compounds such as proteins, water molecules and ions due to their structure: a bilayer of lipid molecules in which the polar head groups face outwards (towards the exterior and interior of the cell, respectively) and the nonpolar tails face inwards at the center of the bilayer, forming a hydrophobic core (for a general review of membrane structure and function, see Gennis, R.B. (1989) *Biomembranes, Molecular Structure and*

Function, Springer: Heidelberg). This barrier enables cells to maintain a relatively higher concentration of desired compounds and a relatively lower concentration of undesired compounds than are contained within the surrounding medium, since the diffusion of these compounds is effectively blocked by the membrane.

5 However, the membrane also presents an effective barrier to the import of desired compounds and the export of waste molecules. To overcome this difficulty, cellular membranes incorporate many kinds of transporter proteins which are able to facilitate the transmembrane transport of different kinds of compounds. There are two general classes of these transport proteins: pores or channels and transporters. The former are  
10 integral membrane proteins, sometimes complexes of proteins, which form a regulated hole through the membrane. This regulation, or 'gating' is generally specific to the molecules to be transported by the pore or channel, rendering these transmembrane constructs selectively permeable to a specific class of substrates; for example, a potassium channel is constructed such that only ions having a like charge and size to that  
15 of potassium may pass through. Channel and pore proteins tend to have discrete hydrophobic and hydrophilic domains, such that the hydrophobic face of the protein may associate with the interior of the membrane while the hydrophilic face lines the interior of the channel, thus providing a sheltered hydrophilic environment through which the selected hydrophilic molecule may pass. Many such pores/channels are  
20 known in the art, including those for potassium, calcium, sodium, and chloride ions.

This pore and channel-mediated system of facilitated diffusion is limited to very small molecules, such as ions, because pores or channels large enough to permit the passage of whole proteins by facilitated diffusion would be unable to prevent the passage of smaller hydrophilic molecules as well. Transport of molecules by this process  
25 is sometimes termed 'facilitated diffusion' since the driving force of a concentration gradient is required for the transport to occur. Permeases also permit facilitated diffusion of larger molecules, such as glucose or other sugars, into the cell when the concentration of these molecules on one side of the membrane is greater than that on the other (also called 'uniport'). In contrast to pores or channels, these integral membrane  
30 proteins (often having between 6-14 membrane-spanning  $\alpha$ -helices) do not form open channels through the membrane, but rather bind to the target molecule at the surface of the membrane and then undergo a conformational shift such that the target molecule is released on the opposite side of the membrane.

However, cells frequently require the import or export of molecules against the  
35 existing concentration gradient ('active transport'), a situation in which facilitated diffusion cannot occur. There are two general mechanisms used by cells for such membrane transport: symport or antiport, and energy-coupled transport such as that

mediated by the ABC transporters. Symport and antiport systems couple the movement of two different molecules across the membrane (via permeases having two separate binding sites for the two different molecules); in symport, both molecules are transported in the same direction, while in antiport, one molecule is imported while the other is exported. This is possible energetically because one of the two molecules moves in accordance with a concentration gradient, and this energetically favorable event is permitted only upon concomitant movement of a desired compound against the prevailing concentration gradient. Single molecules may be transported across the membrane against the concentration gradient in an energy-driven process, such as that utilized by the ABC transporters. In this system, the transport protein located in the membrane has an ATP-binding cassette; upon binding of the target molecule, the ATP is converted to ADP + Pi, and the resulting release of energy is used to drive the movement of the target molecule to the opposite face of the membrane, facilitated by the transporter. For more detailed descriptions of all of these transport systems, see:

15 Bamberg, E. et al., (1993) "Charge transport of ion pumps on lipid bilayer membranes", *Q. Rev. Biophys.* 26: 1-25; Findlay, J.B.C. (1991) "Structure and function in membrane transport systems", *Curr Opin. Struct. Biol.* 1:804-810; Higgins, C.F. (1992) "ABC transporters from microorganisms to man", *Ann. Rev. Cell Biol.* 8: 67-113; Gennis, R.B. (1989) "Pores, Channels and Transporters", in: *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg, p. 270-322; and Nikaido, H. and Saier, H. (1992) "Transport proteins in bacteria: common themes in their design", *Science* 258: 936-942, and references contained within each of these references.

The synthesis of membranes is a well-characterized process involving a number of components, the most important of which are lipid molecules. Lipid synthesis may be divided into two parts: the synthesis of fatty acids and their attachment to sn-glycerol-3-phosphate, and the addition or modification of a polar head group. Typical lipids utilized in bacterial membranes include phospholipids, glycolipids, sphingolipids, and phosphoglycerides. Fatty acid synthesis begins with the conversion of acetyl CoA either to malonyl CoA by acetyl CoA carboxylase, or to acetyl-ACP by acetyltransacylase. Following a condensation reaction, these two product molecules together form acetoacetyl-ACP, which is converted by a series of condensation, reduction and dehydration reactions to yield a saturated fatty acid molecule having a desired chain length. The production of unsaturated fatty acids from such molecules is catalyzed by specific desaturases either aerobically, with the help of molecular oxygen, or anaerobically (for reference on fatty acid synthesis, see F.C. Neidhardt et al. (1996) *E. coli* and *Salmonella*. ASM Press: Washington, D.C., p. 612-636 and references contained therein; Lengeler et al. (eds) (1999) *Biology of Prokaryotes*. Thieme:

Stuttgart, New York, and references contained therein; and Magnuson, K. et al., (1993) *Microbiological Reviews* 57: 522-542, and references contained therein). The cyclopropane fatty acids (CFA) are synthesized by a specific CFA-synthase using SAM as a cosubstrate. Branched chain fatty acids are synthesized from branched chain amino acids that are deaminated to yield branched chain 2-oxo-acids (see Lengeler et al., eds. (1999) *Biology of Prokaryotes*. Thieme: Stuttgart, New York, and references contained therein). Another essential step in lipid synthesis is the transfer of fatty acids onto the polar head groups by, for example, glycerol-phosphate-acyltransferases. The combination of various precursor molecules and biosynthetic enzymes results in the production of different fatty acid molecules, which has a profound effect on the composition of the membrane.

### III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MCT nucleic acid and protein molecules, which control the production of cellular membranes in *C. glutamicum* and govern the movement of molecules across such membranes. In one embodiment, the MCT molecules participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In a preferred embodiment, the activity of the MCT molecules of the present invention to regulate membrane component production and membrane transport has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MCT molecules of the invention are modulated in activity, such that the *C. glutamicum* metabolic pathways which the MCT proteins of the invention regulate are modulated in yield, production, and/or efficiency of production and the transport of compounds through the membranes is altered in efficiency, which either directly or indirectly modulates the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "MCT protein" or "MCT polypeptide" includes proteins which participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of MCT proteins include those encoded by the MCT genes set forth in Table 1 and Appendix A. The terms "MCT gene" or "MCT nucleic acid sequence" include nucleic acid sequences encoding an MCT protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MCT genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the

desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

In another embodiment, the MCT molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursor, cofactor, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the

degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

5 The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, MCT proteins of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from *C. glutamicum* in large-scale fermentative culture.

25 The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MCT cDNAs and the predicted amino acid sequences of the *C. glutamicum* MCT proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins involved in the metabolism of cellular membrane components or proteins involved in the transport of compounds across such membranes.

35 The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein



which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The MCT protein or a biologically active portion or fragment thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

#### A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode MCT polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of MCT-encoding nucleic acid (e.g., MCT DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MCT nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* MCT cDNA can be isolated from a *C. glutamicum* library

5 using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences

10 of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g.,

15 by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can

20 be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore,

25 oligonucleotides corresponding to an MCT nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* MCT cDNAs of the

30 invention. This cDNA comprises sequences encoding MCT proteins (i.e., "the coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

35 For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA number having the designation "RXA" followed by 5 digits (i.e., RXA00001). Each of these sequences

comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequence in Appendix B designated RXA00001 is a translation of the coding region of the nucleotide sequence of nucleic acid molecule RXA00001 in Appendix A.

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the *dapD* gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MCT protein. The nucleotide sequences determined from the cloning of the MCT genes from *C. glutamicum* allows for the generation of probes and

- primers designed for use in identifying and/or cloning MCT homologues in other cell types and organisms, as well as MCT homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone MCT homologues.
- 5 Probes based on the MCT nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MCT protein, such as by measuring a level of an MCT-encoding nucleic acid in a sample of cells, e.g., detecting MCT mRNA levels or determining whether a genomic MCT gene has been mutated or deleted.
- 10 15

- In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Protein members of such membrane component metabolic pathways or membrane transport systems, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an MCT protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MCT protein activities are set forth in Table 1.
- 20 25 30 35

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

5        Portions of proteins encoded by the MCT nucleic acid molecules of the invention are preferably biologically active portions of one of the MCT proteins. As used herein, the term "biologically active portion of an MCT protein" is intended to include a portion, e.g., a domain/motif, of an MCT protein that participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in  
10   the transport of molecules across these membranes, or has an activity as set forth in Table 1. To determine whether an MCT protein or a biologically active portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, an assay of enzymatic activity may be performed. Such assay methods are  
15   well known to those skilled in the art, as detailed in Example 8 of the Exemplification.

      Additional nucleic acid fragments encoding biologically active portions of an MCT protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the MCT protein or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MCT protein  
20   or peptide.

      The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same MCT protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic  
25   acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

30        In addition to the *C. glutamicum* MCT nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MCT proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the MCT gene may exist among individuals within a population due to natural variation. As  
35   used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an MCT protein, preferably a *C. glutamicum* MCT protein. Such natural variations can typically result in 1-5% variance

in the nucleotide sequence of the MCT gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in MCT that are the result of natural variation and that do not alter the functional activity of MCT proteins are intended to be within the scope of the invention.

5 Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MCT cDNA of the invention can be isolated based on their homology to the *C. glutamicum* MCT nucleic acid disclosed herein using the *C. glutamicum* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in  
10 another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for  
15 hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art  
20 and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions  
25 to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MCT protein.

30 In addition to naturally-occurring variants of the MCT sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded MCT protein, without altering the functional ability of the MCT protein. For example, nucleotide substitutions leading to amino acid  
35 substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MCT proteins (Appendix B) without altering the

activity of said MCT protein, whereas an "essential" amino acid residue is required for MCT protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having MCT activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MCT activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MCT proteins that contain changes in amino acid residues that are not essential for MCT activity. Such MCT proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the MCT activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an MCT protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the

09.07.99

encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an MCT protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MCT coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MCT activity described herein to identify mutants that retain MCT activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding MCT proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MCT coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MCT protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID RXA00001 comprises nucleotides 1 to 1128). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MCT. The term "noncoding region" refers to 5' and 3' sequences



which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding MCT disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MCT mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MCT mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MCT mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MCT protein to thereby inhibit expression of the

protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-O-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MCT mRNA transcripts to thereby inhibit translation of MCT mRNA. A ribozyme having specificity for an MCT-encoding nucleic acid can be designed based upon the nucleotide sequence of an MCT cDNA disclosed herein (i.e., RXA00001 in Appendix A). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MCT-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, MCT mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, MCT gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MCT nucleotide sequence (e.g., an MCT promoter and/or enhancers) to form triple helical structures that prevent transcription of an MCT gene in target cells. See generally, Helene, C. (1991)

Anticancer Drug Des. 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

#### B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MCT protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and

09.07.99

those which direct expression of the nucleotide sequence only in certain host cells. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., MCT proteins, mutant forms of MCT proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of MCT proteins in prokaryotic or eukaryotic cells. For example, MCT genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: *More Gene Manipulations in Fungi*, J.W. Benner & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* -mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

09-07-99

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the MCT protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant MCT protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$  prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the MCT protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector

development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

Alternatively, the MCT proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the MCT proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddie (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and

European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine *hox* promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

5 The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to MCT mRNA. Regulatory sequences operatively  
10 linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid  
15 or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., *Antisense RNA as a molecular tool for genetic analysis, Reviews - Trends in Genetics*, Vol. 1(1) 1986.

20 Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due  
25 to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an MCT protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or  
30 mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via  
35 conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g.,

09.07.99

DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an MCT protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an MCT gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the MCT gene. Preferably, this MCT gene is a *Corynebacterium glutamicum* MCT gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous MCT gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous MCT gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous MCT protein). In the homologous recombination vector, the altered portion of the MCT gene is flanked at its 5' and 3' ends by additional nucleic acid of the MCT gene to allow for homologous recombination to occur between the exogenous MCT gene carried by the vector and an endogenous MCT gene in a microorganism. The additional flanking MCT nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and



cells in which the introduced MCT gene has homologously recombined with the endogenous MCT gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene.

- 5 For example, inclusion of an MCT gene on a vector placing it under control of the lac operon permits expression of the MCT gene only in the presence of IPTG. Such regulatory systems are well known in the art.

- A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an MCT protein. Accordingly, the  
10 invention further provides methods for producing MCT proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an MCT protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered MCT protein) in a suitable medium until MCT protein is produced. In another  
15 embodiment, the method further comprises isolating MCT proteins from the medium or the host cell.

### *C. Isolated MCT Proteins*

- Another aspect of the invention pertains to isolated MCT proteins, and  
20 biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of MCT protein in which the protein is separated from cellular components  
25 of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of MCT protein having less than about 30% (by dry weight) of non-MCT protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-MCT protein, still more preferably less than about 10% of non-MCT protein, and most  
30 preferably less than about 5% non-MCT protein. When the MCT protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other  
35 chemicals" includes preparations of MCT protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or

other chemicals" includes preparations of MCT protein having less than about 30% (by dry weight) of chemical precursors or non-MCT chemicals, more preferably less than about 20% chemical precursors or non-MCT chemicals, still more preferably less than about 10% chemical precursors or non-MCT chemicals, and most preferably less than about 5% chemical precursors or non-MCT chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the MCT protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* MCT protein in a microorganism such as *C. glutamicum*

An isolated MCT protein or a portion thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an MCT protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences of Appendix B. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein. For example, a preferred MCT protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 1.

In other embodiments, the MCT protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of

the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the MCT protein is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the MCT activities described herein. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an MCT protein include peptides comprising amino acid sequences derived from the amino acid sequence of an MCT protein, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an MCT protein, which include fewer amino acids than a full length MCT protein or the full length protein which is homologous to an MCT protein, and exhibit at least one activity of an MCT protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an MCT protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an MCT protein include one or more selected domains/motifs or portions thereof having biological activity.

MCT proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the MCT protein is expressed in the host cell. The MCT protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an MCT protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native MCT protein can be isolated from cells (e.g., endothelial cells), for example using an anti-MCT antibody, which can be produced by standard techniques utilizing an MCT protein or fragment thereof of this invention.

The invention also provides MCT chimeric or fusion proteins. As used herein, an MCT "chimeric protein" or "fusion protein" comprises an MCT polypeptide operatively linked to a non-MCT polypeptide. An "MCT polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an MCT protein, whereas a "non-MCT polypeptide" refers to a polypeptide having an amino acid sequence

09-07-99

corresponding to a protein which is not substantially homologous to the MCT protein, e.g., a protein which is different from the MCT protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the MCT polypeptide and the non-MCT polypeptide are fused in-frame to each other. The non-MCT polypeptide can be fused to the N-terminus or C-terminus of the MCT polypeptide. For example, in one embodiment the fusion protein is a GST-MCT fusion protein in which the MCT sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant MCT proteins. In another embodiment, the fusion protein is an MCT protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an MCT protein can be increased through use of a heterologous signal sequence.

Preferably, an MCT chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An MCT-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MCT protein.

Homologues of the MCT protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the MCT protein. As used herein, the term "homologue" refers to a variant form of the MCT protein which acts as an agonist or antagonist of the activity of the MCT protein. An agonist of the MCT protein can retain substantially the same, or a subset, of the biological activities of the MCT protein. An antagonist of the MCT protein can inhibit one or more of the activities of the naturally occurring form of the MCT protein, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the MCT

protein, or by binding to an MCT protein which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

In an alternative embodiment, homologues of the MCT protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the MCT protein for MCT protein agonist or antagonist activity. In one embodiment, a variegated library of MCT variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MCT variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MCT sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of MCT sequences therein. There are a variety of methods which can be used to produce libraries of potential MCT homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential MCT sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the MCT protein coding can be used to generate a variegated population of MCT fragments for screening and subsequent selection of homologues of an MCT protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MCT coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the MCT protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of MCT homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the

gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique  
5 which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MCT homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

10 In another embodiment, cell based assays can be exploited to analyze a variegated MCT library, using methods well known in the art.

#### *D. Uses and Methods of the Invention*

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the  
15 following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of MCT protein regions required for function; modulation of an MCT protein activity; modulation of the metabolism of one or more cell membrane components; modulation of the  
20 transmembrane transport of one or more compounds; and modulation of cellular production of a desired compound, such as a fine chemical.

The MCT nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum*  
25 or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present.  
30 Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. Detection of such organisms is of significant clinical relevance.

Further, the nucleic acid and protein molecules of the invention may serve as  
35 markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated

with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The MCT nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the MCT nucleic acid molecules of the invention may result in the production of MCT proteins having functional differences from the wild-type MCT proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Recovery of fine chemical compounds from large-scale cultures of *C. glutamicum* is significantly improved if *C. glutamicum* secretes the desired compounds, since such compounds may be readily purified from the culture medium (as opposed to extracted from the mass of *C. glutamicum* cells). By either increasing the number or the activity of transporter molecules which export fine chemicals from the cell, it may be possible to increase the amount of the produced fine chemical which is present in the extracellular medium, thus permitting greater ease of harvesting and purification. Conversely, in order to efficiently overproduce one or more fine chemicals, increased amounts of the cofactors, precursor molecules, and intermediate compounds for the appropriate biosynthetic pathways are required. Therefore, by increasing the number and/or activity

of transporter proteins involved in the import of nutrients, such as carbon sources (i.e., sugars), nitrogen sources (i.e., amino acids, ammonium salts), phosphate, and sulfur, it may be possible to improve the production of a fine chemical, due to the removal of any nutrient supply limitations on the biosynthetic process. Further, fatty acids and lipids are themselves desirable fine chemicals, so by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

The engineering of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, the normal biochemical processes of metabolism result in the production of a variety of waste products (e.g., hydrogen peroxide and other reactive oxygen species) which may actively interfere with these same metabolic processes (for example, peroxynitrite is known to nitrate tyrosine side chains, thereby inactivating some enzymes having tyrosine in the active site (Groves, J.T. (1999) *Curr. Opin. Chem. Biol.* 3(2): 226-235). While these waste products are typically excreted, the *C. glutamicum* strains utilized for large-scale fermentative production are optimized for the overproduction of one or more fine chemicals, and thus may produce more waste products than is typical for a wild-type *C. glutamicum*. By optimizing the activity of one or more MCT proteins of the invention which are involved in the export of waste molecules, it may be possible to improve the viability of the cell and to maintain efficient metabolic activity. Also, the presence of high intracellular levels of the desired fine chemical may actually be toxic to the cell, so by increasing the ability of the cell to secrete these compounds, one may improve the viability of the cell.

Further, the MCT proteins of the invention may be manipulated such that the relative amounts of various lipid and fatty acid molecules produced are altered. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, which, as previously explicated, may modify the export of waste products or the produced fine chemical or the import of necessary nutrients. Such membrane fluidity changes may also profoundly affect the integrity of the cell; cells with relatively weaker membranes are more vulnerable in the large-scale fermentor environment to mechanical stresses which may



5 damage or kill the cell. By manipulating MCT proteins involved in the production of fatty acids and lipids for membrane construction such that the resulting membrane has a membrane composition more amenable to the environmental conditions extant in the cultures utilized to produce fine chemicals, a greater proportion of the *C. glutamicum* cells should survive and multiply. Greater numbers of *C. glutamicum* cells in a culture should translate into greater yields, production, or efficiency of production of the fine chemical from the culture.

10 The aforementioned mutagenesis strategies for MCT proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated MCT nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is  
15 improved. This desired compound may be any natural product of *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

### Exemplification

#### **Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032**

10 A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-1 (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-1: 140.34 g/l sucrose,  
15 2.46 g/l  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 10 ml/l  $\text{KH}_2\text{PO}_4$  solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l  $(\text{NH}_4)_2\text{SO}_4$ , 1 g/l NaCl, 2 g/l  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.2 g/l  $\text{CaCl}_2$ , 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l  $\text{FeSO}_4 \cdot \text{H}_2\text{O}$ , 10 mg/l  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 3 mg/l  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 30 mg/l  $\text{H}_3\text{BO}_3$ , 20 mg/l  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ , 1 mg/l  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 3 mg/l  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 500 mg/l complexing agent (EDTA or citric acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting  
25 protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-1 and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca. 18 h at 37°C. The DNA was purified by  
30 extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20

µg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

**Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.**

Starting from DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (see e.g., Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) Proc. Natl. Acad. Sci. USA, 75:3737-3741); pACYC177 (Change & Cohen (1978) J. Bacteriol 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) Gene 53:283-286.

**Example 3: DNA Sequencing and Computational Functional Analysis**

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

**Example 4: *In vivo* Mutagenesis**

*In vivo* mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain

the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is  
5 illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

**Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum***

Several *Corynebacterium* and *Brevibacterium* species contain endogenous  
10 plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in  
15 Molecular Biology", John Wiley & Sons) to which a origin of replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903  
20 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli* and *C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. *et al.* (1985) *J. Bacteriol.* 162:591-597,  
25 Martin J.F. *et al.* (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. *et al.* (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by  
30 protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159:306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A *et al.* (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for

*C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an *Mcr*-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) J. Mol. Biol. 166:1-19).

5

#### **Example 6: Assessment of the Expression of the Mutant Protein**

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

30

### Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for *Corynebacteria* are both well-known and readily available (Lieb *et al.* (1989) Appl. Microbiol. Biotechnol., 32:205-210; von der Osten *et al.* (1998) Biotechnology Letters, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Prokaryotes, Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as  $\text{NH}_4\text{Cl}$  or  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{NH}_4\text{OH}$ , nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology. A Practical Approach (eds. P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0

19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFC) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if  
5 necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to  
10 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH<sub>4</sub>OH during growth. If complex medium components such as yeast extract are utilized,  
15 the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the microorganisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the  
20 broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth  
25 medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control  
30 clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD<sub>600</sub> of 0.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2.5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract,

22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

5 **Example 8 – *In vitro* Analysis of the Function of Mutant Proteins**

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) *Enzymes*. Longmans: London; Fersht, (1985) *Enzyme Structure and Mechanism*. Freeman: New York; Walsh, (1979) *Enzymatic Reaction Mechanisms*. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) *Fundamentals of Enzymology*. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) *The Enzymes*, 3<sup>rd</sup> ed. Academic Press: New York; Bisswanger, H., (1994) *Enzymkinetik*, 2<sup>nd</sup> ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) *Methods of Enzymatic Analysis*, 3<sup>rd</sup> ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's *Encyclopedia of Industrial Chemistry* (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

30 **Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product**

35 The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing



the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

#### Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. *Biochemical Engineering Fundamentals*, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biorekhnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, John Wiley and Sons; Fallon, A. et al. (1987) *Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 17.

### Equivalents

Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

# TABLE 1: GENES IN THE APPLICATION

## Fatty acid and lipid synthesis

Identification Code	Config.	NT Start	NT Stop	Gene Name	Function
RXA02335	GR00872	550	2322	EC-accC	BIOTIN CARBOXYLASE (EC 6.3.4.14)
RXA02173	GR00841	7473	8974	BS-yhlI, EC-accD	ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT BETA (EC 6.4.1.2)
RXA01784	GR00500	2178	3110	BS-labG, EC-labG	3-OXOACYL-JACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
RXA02487	GR00716	4937	4650		LONG-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.3)
RXA02490	GR00720	817	5		LONG-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.3)
RXA01467	GR00422	920	1210		ACYL CARRIER PROTEIN
RXA00798	GR00212	202	5		Acyl carrier protein phosphodiesterase
RXA01897	GR00544	617	1159		Acyl carrier protein phosphodiesterase
RXA02809	GR00790	277	5		Acyl carrier protein phosphodiesterase
RXA00113	GR00017	2	3295		FATTY-ACID SYNTHASE (EC 2.3.1.85)
RXA00158	GR00024	2088	4		FATTY ACID SYNTHASE (EC 2.3.1.85)
RXA02572	GR00155	2	3832		FATTY ACID SYNTHASE (EC 2.3.1.85)
RXA02582	GR00741	1890	6719	BS-pksP, EC-labF	PROBABLE POLYKETIDE SYNTHASE CY338 20
RXA02681	GR00754	15347	14541	BS-yvoA, EC-labR	FATTY ACYL RESPONSIVE REGULATOR
RXA00880	GR00242	6213	8057		LONG-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.3)
RXA01080	GR00296	9568	10489		OMEGA-3 FATTY ACID DESATURASE (EC 1.14.99.1)
RXA01722	GR00488	5746	4022		MEDIUM-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.1)
RXA01644	GR00456	9854	8577	EC-dla	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1.78)
RXA02029	GR00618	355	1669		CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1.78)
RXA01801	GR00509	3396	2380		ENOYL-COA HYDRATASE (EC 4.2.1.17)
RXA02512	GR00721	3303	4259		LIPID A BIOSYNTHESIS LAUROYL ACYL TRANSFERASE (EC 2.3.1.1)
RXA00899	GR00245	1599	2864		CARDIOLIPIN SYNTHETASE (EC 2.7.8.1)
RXA00819	GR00221	18	1007	EC-cls, BS-ywlE	ACYL-COA DEHYDROGENASE (EC 1.3.99.1)
RXA01766	GR00500	4081	4371	BS-yngJ, EC-caiA	ACYL-COA DEHYDROGENASE (EC 1.3.99.1)
RXA01762	GR00500	1272	10	BS-yhlI, EC-ladD	LONG-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.3)
RXA00681	GR00179	3405	2662		3-OXOACYL-JACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
RXA00802	GR00214	3803	4516		3-OXOACYL-JACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
RXA02133	GR00639	3	308	EC-ydlG	3-OXOACYL-JACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
RXA01114	GR00308	2	793	EC-b1397, BS-ywlK	3-KETOACYL-COA THIOLEASE (EC 2.3.1.16)
RXA02810	GR00791	348	4		FATTY ACYL RESPONSIVE REGULATOR
RXA01894	GR00542	1622	2476	EC-cdsA, BS-cdsA	PHOSPHATIDATE CYTIDYL TRANSFERASE (EC 2.7.7.41)
RXA02599	GR00742	3179	3655		PHOSPHATIDYLGLYCEROPHOSPHATASE B (EC 3.1.3.27)
RXA02638	GR00749	8	511		1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYL TRANSFERASE (EC 2.3.1.51)

Identification Code	Config.	NT Start	NT Stop	Gene Name	Function
RXA00856	GR00232	720	1256	BS-pgsA	CDP-DIACYLGLYCEROL-GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYL TRANSFERASE (EC 2.7.8.5)
RXA02511	GR00721	2621	3277		CDP-DIACYLGLYCEROL-GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYL TRANSFERASE (EC 2.7.8.5)
RXA02836	GR00827	106	411	BS-yjgO	KETOACYL REDUCTASE HETN (EC 1.3.1.-)
RXA02578	GR00740	2436	3541		PUTATIVE ACYL TRANSFERASE
RXA02150	GR00639	1856	19558		1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYL TRANSFERASE (EC 2.3.1.51)
RXA00907	GR00160	1869	2249		POLY(3-HYDROXYALKANOATE) POLYMERASE (EC 2.3.1.-)
RXA02397	GR00698	1688	2683		POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.-)
RXA00560	GR00171	1027	5		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA00801	GR00214	3138	3770		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA00821	GR00221	1469	2311		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA01833	GR00517	1666	260	BS-yjgH	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA01853	GR00525	5561	5010	EC-b0827	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA02424	GR00706	808	428		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
RXA00419	GR00095	3	484		ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
RXA00421	GR00096	565	723		ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)

## Polyketide Synthesis

Identification Code	Config.	NT Start	NT Stop	Gene Name	Function
RXA01420	GR00416	775	17		4"-MYCAROSYL ISOVALERYL COA TRANSFERASE (EC 2....)
RXA02581	GR00741	1	1527		POLYKETIDE SYNTHASE
RXA02582	GR00741	1890	6719	BS-phsP, EC-tabf	PROBABLE POLYKETIDE SYNTHASE CY338.20
RXA01138	GR00318	1656	2072		ACTINORHODIN POLYKETIDE DIMERASE (EC....)
RXA01980	GR00573	1470	838		POLYKETIDE CYCLASE

## Fatty acid degradation

Identification Code	Config.	NT Start	NT Stop	Gene Name	Function
RXA02268	GR00655	2182	3081		LIPASE (EC 3.1.1.3)
RXA02269	GR00655	3094	4065		LIPASE (EC 3.1.1.3)
RXA01614	GR00449	8219	7197		LYSOPHOSPHOLIPASE L2 (EC 3.1.1.5)
RXA01983	GR00573	3559	3053		LIPASE (EC 3.1.1.3)

Identification Code	NT	Start	Contig.	NT	Slop	Gene Name	Function
RXA02320		593	GR00667		6		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02321		1380	GR00667		937		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02343		1403	GR00675		1816		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02583		6743	GR00741		8290	BS-yqjD	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02850		2	GR00850		493		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02851		524	GR00851		6		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA00870		809	GR00239		2320	BS-lolA	METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING) (EC 1.2.1.27)
							2-Methyl-3-oxopropanoate NAD <sup>+</sup> oxidoreductase (CoA-propanoylating)
RXA01260		2381	GR00367		1200	BS-pdhD, EC-lpdA	LIPAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
RXA01261		2607	GR00367		2437		LIPAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
RXA00831		4859	GR00253		4114	EC-tesB	ISOVALERYL-COA DEHYDROGENASE (EC 1.3.99.10)
RXA01136		685	GR00318		1116		ISOVALERYL-COA DEHYDROGENASE (EC 1.3.99.10)
RXA00558		218	GR00149		8		PROTEIN VLDL
RXA01580		707	GR00440		6		Glycerophosphoryl diester phosphodiesterase
RXA02677		3119	GR00754		3877	BS-yhdW	GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASE (EC 3.1.4.48)

### Lipoprotein and Lipopolysaccharide synthesis

Identification Code	NT	Start	Contig.	NT	Slop	Gene Name	Function
RXA00002		2278	GR00001		1595		DOLICHOYL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2.4.1.83) /
RXA00180		4044	GR00024		4616		APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.)
RXA00345		80	GR00084		1040		(LIPOPROTEIN NLPD) PPB HOMOLOG PRECURSOR
RXA00413		3859	GR00092		2863	EC-yaeC, BS-yhcJ	Zn-binding lipoprotein
RXA00482		18891	GR00119		18244		OUTER MEMBRANE LIPOPROTEIN 3 PRECURSOR
RXA01164		1579	GR00332		5	EC-lml	OUTER MEMBRANE LIPOPROTEIN BLC PRECURSOR
RXA01168		1285	GR00333		566		DOLICHOYL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2.4.1.83) /
RXA02087		3159	GR00826		1990		APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.)
RXA02222		9420	GR00651		9794		DOLICHOYL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2.4.1.83) /
RXA02313		5812	GR00665		4592		APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.)
RXA02491		902	GR00720		2155		DOLICHOYL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2.4.1.83) /
RXA02585		19052	GR00741		19702	BS-ybN	APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.)
RXA02616		598	GR00745		1308		Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02627		2981	GR00747		2139		Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02650		1460	GR00752		2038	EC-nlpD	Lipopolysaccharide N-acetylglucosaminyltransferase
							LIPOPROTEIN NLPD PRECURSOR
							DTX/IRON-REGULATED LIPOPROTEIN PRECURSOR
							LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36)

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01094	GR00308	2703	1758		PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-)
RXA01985	GR00574	466	1008		DTXR/IRON-REGULATED LIPOPROTEIN PRECURSOR
RXA02804	GR00785	1	261		DTXR/IRON-REGULATED LIPOPROTEIN PRECURSOR
RXA00934	GR00253	6835	6047		(AE000805) LPS biosynthesis RhuU related protein [Methanobacterium thermoautotrophicum]
RXA02805	GR00742	11557	12051		ANTIGEN 85-B PRECURSOR

## Terpenoid biosynthesis

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00875	GR00241	2423	1857	EC-b7889	ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (EC 5.3.3.2)
RXA01292	GR00373	1204	2388		PHYTOENE DEHYDROGENASE (EC 1.3.-.)
RXA01293	GR00373	2370	2686		PHYTOENE DEHYDROGENASE (EC 1.3.-.)
RXA02310	GR00665	1132	2394		GERANYLGERANYL HYDROGENASE
RXA02718	GR00758	18539	19585		GERANYLGERANYL PYROPHOSPHATE SYNTHASE (EC 2.5.1.1)
RXA01067	GR00298	1453	2181	EC-b0174	undecaprenyl-diphosphate synthase (EC 2.5.1.31)
RXA01269	GR00367	20334	18894	BS-yvC, EC-b7047	UNDECAPRENYL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE (EC 2.7.8.6)
RXA01205	GR00346	3	533		PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
(EC 2.4.1.-)					
RXA01576	GR00438	8053	8811		DOLICHYL-PHOSPHATE BETA-GLUCOSYLTRANSFERASE (EC 2.4.1.117)
RXA01258	GR00366	1150	2709		DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSYLTRANSFERASE 1 (EC 2.4.1.109)
RXA01351	GR00392	2841	3137	BS-yvdD	(U15180) P450 cytochrome, isopenicillin, ferrisox [Mycobacterium leprae]
RXA02309	GR00665	978	4	BS-gerCC, EC-ispB	OCTAPRENYL-DIPHOSPHATE SYNTHASE (EC 2.5.1.-)
RXA00477	GR00119	13187	11544		PHYTOENE DEHYDROGENASE (EC 1.3.-.)
RXA00478	GR00119	14020	13190		PHYTOENE SYNTHASE (EC 2.5.1.-)
RXA01291	GR00373	345	1277		PHYTOENE SYNTHASE (EC 2.5.1.-)
RXA00480	GR00119	17444	16329		FARNESYL DIPHOSPHATE SYNTHASE (EC 2.5.1.1) (EC 2.5.1.10)

## ABC-Transporter

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01946	GR00559	1849	575		(AL021184) ABC transporter ATP binding protein [Mycobacterium tuberculosis]
RXA00164	GR00075	1782	94	EC-msbA	P, G, R ATPase subunits of ABC transporters
RXA00165	GR00075	3275	1860		P, G, R ATPase subunits of ABC transporters
RXA00243	GR00037	830	4		P, G, R ATPase subunits of ABC transporters

# Identification

Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00259	GR000339	8469	6268		P, G, R ATPase subunits of ABC transporters
RXA00410	GR000982	829	164		P, G, R ATPase subunits of ABC transporters
RXA00456	GR00114	316	5		P, G, R ATPase subunits of ABC transporters
RXA00459	GR00115	1231	245		P, G, R ATPase subunits of ABC transporters
RXA01604	GR00448	2	607		P, G, R ATPase subunits of ABC transporters
RXA02547	GR00726	22055	19932		P, G, R ATPase subunits of ABC transporters
RXA07571	GR00736	1469	2497		P, G, R ATPase subunits of ABC transporters
RXA07074	GR00628	5798	4176		P, G, R ATPase subunits of ABC transporters
RXA02095	GR00629	14071	15474		P, G, R ATPase subunits of ABC transporters
RXA02225	GR00652	3155	2275		P, G, R ATPase subunits of ABC transporters
RXA07253	GR00654	20480	21406		P, G, R ATPase subunits of ABC transporters
RXA01881	GR00537	3082	3532	BS-yvcJ, EC-b3205	ATPase components of ABC transporters with duplicated ATPase domains
RXA00188	GR00029	1447	5	EC-yjk, BS-ymR	Hypothetical ABC Transporter ATP-Binding Protein
RXA00526	GR00136	1353	664	BS-yknY, EC-b0879	Hypothetical ABC Transporter ATP-Binding Protein
RXA00733	GR00197	411	4		Hypothetical ABC Transporter ATP-Binding Protein
RXA00734	GR00197	863	411		Hypothetical ABC Transporter ATP-Binding Protein
RXA00735	GR00198	849	181		Hypothetical ABC Transporter ATP-Binding Protein
RXA00878	GR00242	3733	1871		Hypothetical ABC Transporter ATP-Binding Protein
RXA01191	GR00341	1571	165	BS-ywA	Hypothetical ABC Transporter ATP-Binding Protein
RXA01212	GR00350	1	813	EC-molB	Hypothetical ABC Transporter ATP-Binding Protein
RXA01847	GR00360	812	534	EC-yadG	Hypothetical ABC Transporter ATP-Binding Protein
RXA02749	GR00764	4153	5028	BS-yil	Hypothetical ABC Transporter ATP-Binding Protein
RXA02724	GR00652	2271	475		Hypothetical ABC Transporter ATP-Binding Protein
RXA00525	GR00136	664	5		Hypothetical ABC Transporter Pemease Protein
RXA00536	GR00146	1	594	BS-yknZ	Hypothetical ABC Transporter Pemease Protein
RXA02750	GR00764	5079	5894		Hypothetical ABC Transporter Pemease Protein
RXA02096	GR00629	15458	16774		Hypothetical ABC Transporter Pemease Protein
RXA01190	GR00340	1102	194		MULTIDRUG RESISTANCE-LIKE ATP-BINDING PROTEIN MDL
RXA01808	GR00509	8993	7875		PUTATIVE ABC TRANSPORTER
RXA02582	GR00737	798	1515	BS-ydbJ	PUTATIVE ABC TRANSPORTER
RXA00950	GR00760	173	1078	BS-ydlf	similar to ABC transporter (ATP-binding protein)
RXA02119	GR00636	4222	2582	BS-lagH	similar to ABC transporter (ATP-binding protein)
RXA00431	GR00099	119	793		ABCA PROTEIN two-component ABC transporter involved in the metabolism of two wall teichoic acids
RXA01185	GR00338	2451	1594		ATP-BINDING PROTEIN
RXA02434	GR00708	4642	3590	BS-yusC, EC-abc	ATP-BINDING PROTEIN
RXA00412	GR00092	2764	1685	BS-yvgX, EC-b0484	ATP-BINDING PROTEIN ABC
RXA01323	GR00385	1175	3439		similar to heavy metal transporting ATPase
RXA00001	GR00001	1386	259	BS-mtmX, EC-ugpC	SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC

# Other transporters

Identification	NT	NT	Contig	Gene Name	Function
Code	Start	Slop			
RXA07402	747	4	GR00700	BS-yacS	(AF027868) putative transporter [Bacillus subtilis]
RXA07626	2012	1008	GR00747	BS-nrgA, EC-amIB	(AF027868) putative transporter [Bacillus subtilis]
RXA072261	30936	32291	GR00654	BS-gabP, EC-gabP	AMMONIUM TRANSPORT SYSTEM
RXA07020	1015	5	GR00613		AROMATIC AMINO ACID TRANSPORT PROTEIN AROP
RXA00281	4721	5404	GR00043		BACITRACIN TRANSPORT ATP-BINDING PROTEIN BCRA
RXA00570	1	498	GR00153		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA00571	2	1186	GR00154		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA00962	2	667	GR00268		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA02811	177	560	GR00792		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA02115	2	1198	GR00635		BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II CARRIER PROTEIN
RXA00590	178	584	GR00157		BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II CARRIER PROTEIN
RXA01538	5040	5429	GR00427	BS-yacG	C4-DICARBOXYLATE TRANSPORT PROTEIN
RXA01727	1471	194	GR00489	BS-bmQ, EC-bmQ	CHROMATE TRANSPORT PROTEIN
RXA00873	6525	7882	GR00163	EC-dcaA, BS-ydbH	COBAL T TRANSPORT ATP-BINDING PROTEIN CBIO
RXA01584	55	597	GR00441		COBAL T TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00702	2165	846	GR00182	BS-yadD	COBAL T TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00828	1687	1319	GR00273		COBAL T TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00852	3137	2448	GR00231	BS-ybaA	COBAL T TRANSPORT PROTEIN CBIO
RXA00877	1213	68	GR00181		COBAL T TRANSPORT PROTEIN CBIO
RXA00851	1319	567	GR00223		COBAL T TRANSPORT PROTEIN CBIO
RXA00939	2448	1840	GR00231		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA01245	1501	1334	GR00256		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA01247	2	1768	GR00360		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA00089	256	489	GR00381	BS-yycB	CYANATE TRANSPORT PROTEIN CYNX
RXA01584	8172	9344	GR00014	EC-b1791	CYANATE TRANSPORT PROTEIN CYNX
RXA00634	746	9	GR00437	EC-b0709	DI-/TRIPEPTIDE TRANSPORTER
RXA02451	3732	5114	GR00166	BS-ydcF, EC-b1834	DI-/TRIPEPTIDE TRANSPORTER
RXA02394	3484	5007	GR00710	EC-yiaN	DICARBOXYLATE TRANSPORTER
RXA01012	1895	585	GR00897	EC-h0829, BS-appf	DIPEPTIDE TRANSPORT ATP-BINDING PROTEIN DPPB
RXA02860	3748	2108	GR00288		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA02861	548	1186	GR00753		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA02034	1239	1457	GR00753		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA01013	1787	822	GR00619		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA02033	4549	3755	GR00288		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
RXA01006	800	12	GR00619		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
RXA02312	862	5	GR00287		DIPEPTIDE TRANSPORTER PROTEIN DPPB
RXA02762	4459	3101	GR00655	EC-cycA, BS-ydgF	D-SERINE/D-ALANINE/GLYCINE TRANSPORTER
RXA00090	346	630	GR00768	BS-ydO	D-XULOSE PROTON-SYMPORTER
RXA00089	6644	7762	GR00013	BS-ydO	FERRIC ANGUIBACTIN TRANSPORT SYSTEM PERMEASE PROTEIN FATC
RXA01285	5656	6654	GR00013	BS-ydN	FERRIC ANGUIBACTIN TRANSPORT SYSTEM PERMEASE PROTEIN FATC
RXA02728	3	545	GR00371		FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
	184	998	GR00761		FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC



Identification	NT	Start	Contig	NT	Slop	Gene Name	Function
RXA02864	2808	GR10007	2027				FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
RXA05053	30	GR00135	778				FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEFG
RXA01289	2376	GR00372	3419			EC-lepD	FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEFG
RXA01280	3412	GR00372	4575			EC-lepG,BS-ythA	FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEFG
RXA01872	6	GR00516	587				FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEFG
RXA00466	947	GR00117	1933				Ferriochrome transport proteins
RXA02863	2000	GR10007	1026				Ferriochrome transport proteins
RXA01986	622	GR00575	5				GALACTOSE-PROTON SYMPORT
RXA02447	1	GR00710	270				GALACTOSE-PROTON SYMPORT
RXA02761	153	GR00768	353				GALACTOSE-PROTON SYMPORT
RXA02768	1	GR00771	711			BS-ydjK,EC-xyE	GALACTOSE-PROTON SYMPORT
RXA00092	1	GR00014	204				GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
RXA02618	1914	GR00745	2351				GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
RXA02780	2979	GR10040	2128			BS-ytsC	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
RXA01591	3	GR00446	947				GLYCINE BETALINE TRANSPORTER BETP
RXA00201	191	GR00032	6				HIGH AFFINITY RIBOSE TRANSPORTER RBSB
RXA01221	2108	GR00354	2833			EC-livG	HIGH AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN BRAG
RXA01227	2844	GR00354	3542			EC-livF	HIGH AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN LIVF
RXA01219	151	GR00354	1032				HIGH AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PERMEASE PROTEIN LIVH
RXA01220	1032	GR00354	2108				HIGH AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PERMEASE PROTEIN LIVM
RXA00091	7762	GR00013	8514			BS-yclP	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00228	29232	GR00032	28642				IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00346	1054	GR00064	1743				IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00524	779	GR00135	1111			BS-yusV,EC-lepC	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA01823	591	GR00516	1367			BS-ylnA	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA02767	1032	GR00770	1814				IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA02792	8581	GR00777	7829				IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA01235	1165	GR00358	194			BS-ytmD	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA01419	888	GR00415	1151				IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
RXA02794	10172	GR00777	9552			EC-lecC	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
RXA02865	3832	GR10007	2816				IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD MATIC AMINO ACID TRANSPORT
PROTEIN AROP							
RXA00123	7029	GR00019	5911				MAGNESIUM AND COBALT TRANSPORT PROTEIN CORA
RXA02354	473	GR00582	1261				MALTOSE TRANSPORT SYSTEM PERMEASE PROTEIN MALG
RXA02441	5940	GR00709	5284				MANGANESE TRANSPORT SYSTEM ATP-BINDING PROTEIN MNTA
RXA02442	5970	GR00709	6818			EC-yebI	MANGANESE TRANSPORT SYSTEM MEMBRANE PROTEIN MNTB
RXA01756	2069	GR00498	762			BS-ykoK	MG2+ TRANSPORTER MGTE
RXA02068	2	GR00627	1120				MG2+ TRANSPORTER MGTE
RXA00565	135	GR00174	572			BS-ypiD	MG2+ICITRATE COMPLEX SECONDARY TRANSPORTER
RXA02808	1	GR00789	258				MG2+ICITRATE COMPLEX SECONDARY TRANSPORTER
RXA00444	626	GR00106	1402			BS-yvgM,EC-cysU	MOLYBDENUM TRANSPORT SYSTEM PERMEASE PROTEIN MODB
RXA02614	5964	GR00743	5236			BS-ygal,EC-ycbE	NITRATE TRANSPORT ATP-BINDING PROTEIN NRTC
RXA01142	721	GR00320	302				NITRATE TRANSPORT ATP-BINDING PROTEIN NRTD
RXA01135	327	GR00318	4				NITRATE TRANSPORT PROTEIN NRTA
RXA01141	636	GR00319	175				NITRATE TRANSPORT PROTEIN NRTA

Identification Code	NT	Start	Stop	Gene Name	Function
RXA00728	1658	2449	NT	BS-yckA,EC-hisM	NOPALINE TRANSPORT SYSTEM PERMEASE PROTEIN NOCM
RXA02795	3	1097			OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN APPF
RXA00761	8530	9120			OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA01939	2042	1440			OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA02863	2059	3453		BS-oppD,EC-oppD	OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA02864	3811	4270			OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPF
RXA00739	6580	7503		EC-dppB,BS-dppB	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPB
RXA00760	7499	8530		BS-oppC,EC-b0832	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPC
RXA02035	3295	1787		EC-b0830	PERIPLASMIC DIPEPTIDE TRANSPORT PROTEIN PRECURSOR
RXA00775	6057	5287		EC-psIB,BS-yqgK	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PSTB
RXA00776	7016	6096		EC-psIA,BS-yqgI	PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN PSTA
RXA00777	8098	7034		EC-psIC,BS-yqgH	PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN PSTC
RXA00774	4546	5199			PHOSPHATE TRANSPORT SYSTEM REGULATORY PROTEIN
RXA01002	3	419		EC-phnC	PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNC
RXA01000	2	541			PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
RXA01003	419	1222			PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
RXA00183	10101	9259		EC-ugpA	POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA01298	1254	882			POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA07422	8200	8634			POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA02515	984	1719		BS-yuiY,EC-b1682	POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA00732	826	5		BS-yfiB	PROBABLE ATP-DEPENDENT TRANSPORTER YCF16
RXA00181	3954	2383		BS-opeU,EC-puIP	PROLINE TRANSPORT SYSTEM
RXA00591	229	1581		EC-b1981	PROLINE/BETAINE TRANSPORTER
RXA01629	3476	1965		EC-pioP	PROLINE/BETAINE TRANSPORTER
RXA02030	3072	1687		EC-yfiE	PROLINE/BETAINE TRANSPORTER
RXA01995	1362	2015		EC-yjiS	PUTATIVE 3-(3-HYDROXYPHENYL) PROPIONATE TRANSPORT PROTEIN
RXA00953	1830	1078			PUTATIVE TRANSPORT PROTEIN SGAT
RXA01188	1585	482			PUTATIVE TRANSPORT PROTEIN SGAT
RXA01972	2116	1523			QUATERNARY AMINE TRANSPORTER
RXA00204	3783	2212		BS-rbsA,EC-rbsA	RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
RXA02438	3236	2478			RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
RXA00203	2152	1241		BS-rbsC,EC-rbsC	RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA00270	2720	1833			RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA02439	4258	3236			RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA00311	1582	738			SHIKIMATE TRANSPORTER
RXA00312	2086	1641			SHIKIMATE TRANSPORTER
RXA01411	1	327			SHIKIMATE TRANSPORTER
RXA01900	2822	4120			SHIKIMATE TRANSPORTER
RXA02507	18760	21160			SHORT-CHAIN FATTY ACIDS TRANSPORTER
RXA00186	12242	12988			SHORT-CHAIN FATTY ACIDS TRANSPORTER
RXA00187	13097	13447		EC-aloB	SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC
RXA00445	21	932			SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC
RXA01880	874	155			SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPA
RXA02353	6	473			SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPE
RXA01287	828	29		BS-yurM,EC-ugpE	SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPE

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01667	GR00464	703	1908		SODIUM/GLUTAMATE SYMPORT CARRIER PROTEIN
RXA02171	GR00641	6571	4919	BS-ywca,EC-yjcG	SODIUM/PROLINE SYMPORTER
RXA00902	GR00245	4643	5875		SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN
RXA00941	GR00257	1999	683	BS-ycae,EC-ydE	sodium-dependent phosphate transport protein
RXA00449	GR00109	2040	1036		Sodium-Dicarboxylate Symport Protein
RXA01755	GR00498	352	5		Sodium-Dicarboxylate Symport Protein
RXA00269	GR00041	1828	1038		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POT
RXA00368	GR00078	1	579		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POT
RXA00389	GR00076	583	1299		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POT
RXA00370	GR00077	6	803		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POT
RXA02073	GR00628	4178	2647		TRANSPORT ATP-BINDING PROTEIN CYDC
RXA01399	GR00409	1	1119	BS-cydc	TRANSPORT ATP-BINDING PROTEIN CYDD
RXA01339	GR00399	8408	7164		TYROSINE-SPECIFIC TRANSPORT PROTEIN
RXA02527	GR00725	5519	6847	BS-yfs,EC-b0770	2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR
RXA00288	GR00048	4459	6303	EC-bet1,BS-opuD	Ecdone/Proline/Glycine betaine carrier ecdp
RXA00586	GR00159	335	787		potassium efflux system protein phae
RXA02044	GR00623	528	1910	BS-ykaB	similar to low-affinity inorganic phosphate transporter
RXA00104	GR00014	15895	16850		CYSO PROTEIN, ammonium transport protein
RXA02384	GR00686	841	215	EC-yibD	C4-DICARBOXYLATE-BINDING PERIPLASMIC PROTEIN PRECURSOR, transport protein
RXA02099	GR00630	6188	6470		AMMONIUM TRANSPORT SYSTEM

## Permeases

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02581	GR00732	664	5	BS-ydcG	NUCLEOSIDE PERMEASE NUPG
RXA02566	GR00733	782	345		NUCLEOSIDE PERMEASE NUPG
RXA00034	GR00004	2	511		Permeases
RXA00190	GR00029	4544	2394		Permeases
RXA00842	GR00228	3208	2009	BS-ykvl	Permeases
RXA01553	GR00432	234	1553		Permeases
RXA00051	GR00008	5770	7173	EC-b0402,BS-ytnA	PROLINE-SPECIFIC PERMEASE PROY
RXA01172	GR00334	2687	4141	BS-ybaR	SULFATE PERMEASE
RXA02128	GR00637	2906	4600		SULFATE PERMEASE
RXA02634	GR00748	6045	7655	BS-yvdB,EC-ychM	SULFATE PERMEASE
RXA02233	GR00653	6856	8067	EC-b1006,BS-pyIP	URACIL PERMEASE
RXA02372	GR00688	6	560		XANTHINE PERMEASE
RXA02377	GR00689	3336	4526	EC-b2888,BS-pbuX	XANTHINE PERMEASE
RXA00081	GR00012	8453	7101		(Predicted) amino acid permeases

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02676	GR00754	2697	1309	EC-yigT,BS-gmlP	GLUCONATE PERMEASE
RXA00432	GR00100	1	691	BS-als1	NAI(+)-LINKED D-ALANINE GLYCINE PERMEASE
RXA00436	GR00101	45	569	EC-yeaJ	NAI(+)-LINKED D-ALANINE GLYCINE PERMEASE
RXA00847	GR00230	1829	381		OLIGOPEPTIDE-BINDING PROTEIN APPA PRECURSOR (permease)
RXA01382	GR00405	1067	6		OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)
RXA02859	GR00753	2	313		OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)

## Channel Proteins

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01395	GR00408	6106	5021	EC-b3001	POTASSIUM CHANNEL BETA SUBUNIT
RXA01737	GR00493	2913	3971		POTASSIUM CHANNEL PROTEIN
RXA00596	GR00159	335	787		potassium efflux system protein phaE
RXA02079	GR00628	9034	9648		CATION EFFLUX SYSTEM PROTEIN CZCD
RXA01303	GR00376	1724	390	EC-naiU	NITRITE EXTRUSION PROTEIN
RXA02079	GR00628	9034	9648		CATION EFFLUX SYSTEM PROTEIN CZCD
RXA00832	GR00224	2239	1685	EC-chaA	CALCIUM/PROTON ANTIPORTER
RXA00378	GR00081	3271	1499		Cation transport ATPases
RXA00842	GR00257	2408	2203		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA01338	GR00389	6984	5087		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA01625	GR00452	3850	3850		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA02220	GR00951	3205	5880	BS-yloB,EC-mgIA	CATION-TRANSPORTING ATPASE PMA1 (EC 3.6.1.-)
RXA00980	GR00276	2648	3286		CATION-TRANSPORTING P-TYPE ATPASE B (EC 3.6.1.-)
RXA02344	GR00876	682	5	EC-kup	KUP SYSTEM POTASSIUM UPTAKE PROTEIN
RXA02348	GR00677	1719	586		KUP SYSTEM POTASSIUM UPTAKE PROTEIN
RXA02476	GR00707	2185	633		PROBABLE NA(+)/H(+) ANTIPORTER
RXA00980	GR00266	583	105		PROTON/SODIUM-GLUTAMATE SYMPORT PROTEIN
RXA01070	GR00299	2089	704	EC-b1728	PROTON/SODIUM-GLUTAMATE SYMPORT PROTEIN
RXA02678	GR00748	6	410		LARGE CONDUCTANCE MECHANOTENSITIVE CHANNEL
RXA01395	GR00408	6106	5021	EC-b3001	POTASSIUM CHANNEL BETA SUBUNIT
RXA01737	GR00493	2913	3971		POTASSIUM CHANNEL PROTEIN

## Other membrane proteins

Identification Code	NT		Gene Name	Function
	Start	Stop		
RXA02597	GR00742	2329	542	OUTER MEMBRANE USHER PROTEIN FIMC PRECURSOR integral membrane protein integral membrane protein MEMBRANE-BOUND PROTEIN LYTR
RXA01454	GR00420	270	4	
RXA01455	GR00420	745	284	
RXA02684	GR00754	8923	8060	
			BS-ywf	

09.07.99

# TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	png	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvate carboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moczek, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 07/20/95
AB003132	murC, flsQ, flsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the flsZ gene from corynebacterium," <i>Biochem Biophys Res Commun</i> , 236(2):383-388 (1997)
AB015023	murC; flsQ		Wachi, M. et al. "A murC gene from Corynebacterium," <i>Appl. Microbiol Biotechnol</i> , 51(2):223-228 (1999)
AB018530	disR		Kimura, E. et al. "Molecular cloning of a novel gene, disR, which rescues the detergent sensitivity of a mutant derived from <i>Brevibacterium lactofermentum</i> ," <i>Biochim Biophys Acta</i> , 1265:156-1570 (1996)
AB018531	disR1; disR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkl	transketolase	
AB024708	glbB, glbD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenylyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	gluA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamoyltransferase	
AF036932	aroD	3-dehydroquinate dehydratase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF038548	pyc	Pyruvate carboxylase	Wehmeier, I. et al. "The role of the <i>Corynebacterium glutamicum</i> rel gene in (p)ppGpp metabolism," <i>Microbiology</i> , 144, 1853-1862 (1998)
AF038651	dcilAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase, ornithine acetyltransferase; N-acetylglutamate kinase, acetylornithine transaminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	Paik, S. et al. "Isolation and analysis of melA, a methionine biosynthetic gene encoding homoserine acetyltransferase in <i>Corynebacterium glutamicum</i> ," <i>Mol Cells</i> , 8(3):286-294 (1998)
AF052652	melA	Homoserine O-acetyltransferase	
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	
			Dusch, N. et al. "Expression of the <i>Corynebacterium glutamicum</i> panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in <i>Escherichia coli</i> ," <i>Appl. Environ. Microbiol.</i> , 65(4):1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinate; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP." <i>J. Bacteriol.</i> , 180(22):6005-6012 (1998)
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Wehrmann, A. et al. "Different modes of diamminopimelate synthesis and their role in cell wall integrity" A study with <i>Corynebacterium glutamicum</i> ." <i>J. Bacteriol.</i> , 180(12):3159-3165 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete)	
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase, $\gamma$ , high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD, strp; amtP	Involved in cell division, PII protein; uridylyltransferase (uridylyl-removing enzyme); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in <i>Corynebacterium glutamicum</i> ; Isolation of genes involved in biochemical characterization of corresponding proteins." <i>FEMS Microbiol.</i> , 173(2):303-310 (1999)
AJ132968	cal	Chloramphenicol acetyl transferase	
AJ224946	mgo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from <i>Corynebacterium glutamicum</i> ." <i>Eur. J. Biochem.</i> , 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of <i>Corynebacterium glutamicum</i> . The channel is formed by a low molecular mass polypeptide." <i>Biochemistry</i> , 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Veres, A.A. et al. "Isolation and characterization of IS31831, a transposable element from <i>Corynebacterium glutamicum</i> ." <i>Mol. Microbiol.</i> , 11(4):739-746 (1994)



GenBank <sup>TM</sup> Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the <i>Corynebacterium glutamicum</i> (Brevibacterium lactofermentum AJ12036) <i>odhA</i> gene encoding a novel type of 2-oxoglutarate dehydrogenase," <i>Microbiology</i> , 142: 3347-3354 (1996)
E01358	hdh, hsk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL, trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al. "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and deshydrobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Deshydrobiotin synthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and deshydrobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al. "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993036782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993036782-A 3 03/09/93
E04484		Prephenate dehydratase	Solouchi, N. et al. "Production of L-phenylalanine by fermentation," Patent: JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N. et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93
E05112		Dihydro-dipicolinate synthetase	Hatakeyama, K. et al. "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M. et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent: JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent: JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent: JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membranous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomercereductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomercereductase," Patent: JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desilhiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in corynebacterium," Patent: JP 1995031476-A 1 02/03/95
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in corynebacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08649		Aspartase	Kohama, K. et al. "DNA fragment having promoter function in <i>Corynebacterium</i> ," Patent. JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al. "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent. JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent. JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-tyrosophan," Patent. JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12764		Arginyl-IRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K. et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent. JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Morckel, B. et al. "Functional and structural analysis of the threonine dehydratase of <i>Corynebacterium glutamicum</i> ," <i>J. Bacteriol.</i> , 174.8065-8072 (1992)
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C. et al. "The cloning and nucleotide sequence of <i>Corynebacterium glutamicum</i> 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," <i>FEBS Microbiol. Lett.</i> , 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetylhydroxy acid synthase large subunit; Acetylhydroxy acid synthase small subunit; Acetylhydroxy acid isomercroductase	Keilhauser, C. et al. "Isoleucine synthesis in <i>Corynebacterium glutamicum</i> . molecular analysis of the ilvB-ilvN-ilvC operon," <i>J. Bacteriol.</i> , 175(17):5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	ptsM	Phosphoenolpyruvate sugar phosphotransferase	Fouci, A et al. "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system" expression in <i>Escherichia coli</i> and homology to enzymes II from enteric bacteria," <i>PNAS USA</i> , 84(24) 8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," <i>FEMS Microbiol Lett</i> , 119(1-2) 137-145 (1994)
L27123	accB	Malate synthase	Lee, H-S. et al. "Molecular characterization of accB, a gene encoding malate synthase in Corynebacterium glutamicum," <i>J Microbiol. Biotechnol.</i> , 4(4) 256-263 (1994)
L27126		Pyruvate kinase	Jellen, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," <i>Appl Environ Microbiol.</i> , 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," <i>J. Bacteriol.</i> , 177(2):465-467 (1995)
L35906	dtxR	Diphtheria toxin repressor	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," <i>J. Bacteriol.</i> , 167:695-702 (1986)
M13774		Prephenate dehydratase	Park, Y-H et al. "Phylogenetic analysis of the coryneform bacteria by 56 rRNA sequences," <i>J Bacteriol.</i> , 169:1801-1806 (1987)
M16175	5S rRNA		Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," <i>Gene</i> , 52:191-200 (1987)
M16663	trpE	Anthraniolate synthase, 5' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," <i>Gene</i> , 52:191-200 (1987)
M16664	trpA	Tryptophan synthase, 3' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," <i>Gene</i> , 52:191-200 (1987)
M23819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," <i>Gene</i> , 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," <i>J. Gen Microbiol.</i> , 138:1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes." <i>J. Gen. Microbiol.</i> , 138, 1167-1173 (1992)
M89931	aecD; bmQ, yhbW	Beta C-S lyase, branched-chain amino acid uptake carrier, hypothetical protein yhbW	Rosol, J. et al. "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminocetyl-L-cysteine." <i>J. Bacteriol.</i> , 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the bmQ gene product." <i>Arch. Microbiol.</i> , 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D. M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence." <i>Appl. Environ. Microbiol.</i> , 59(3):791-799 (1993)
U11545	trpD	Anthraniolate phosphoribosyltransferase	O'Gara, J. P. and Dunican, L. K. (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 trpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR, cglIIR	Putative type II 5-cytosine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli." <i>J. Bacteriol.</i> , 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an MtrBC-deficient Escherichia coli strain." <i>Gene</i> , 203(2):95-101 (1997)
U14965	recA		Ankri, S. et al. "Mutations in the Corynebacterium glutamicum proline biosynthetic pathway: A natural bypass of the proA step." <i>J. Bacteriol.</i> , 178(15):4412-4419 (1996)
U31224	ppx		Ankri, S. et al. "Mutations in the Corynebacterium glutamicum proline biosynthetic pathway: A natural bypass of the proA step." <i>J. Bacteriol.</i> , 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the Corynebacterium glutamicum proline biosynthetic pathway: A natural bypass of the proA step." <i>J. Bacteriol.</i> , 178(15):4412-4419 (1996)
U31230	obg; proB, unkdh	gamma-glutamyl kinase; similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicum proline biosynthetic pathway: A natural bypass of the proA step." <i>J. Bacteriol.</i> , 178(15):4412-4419 (1996)

GenBank <sup>TM</sup> Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of <i>Methylobacillus flagellatum</i> and <i>Corynebacterium glutamicum</i> ," <i>Gene</i> , 175:15-22 (1996)
U35023	thiR, accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," <i>Arch. Microbiol.</i> , 166(2):76-82 (1996)
U43535	cmi	Multidrug resistance protein	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene conferring multidrug resistance in the heterologous host <i>Escherichia coli</i> ," <i>J. Bacteriol.</i> , 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5'-aminoglycoside phosphotransferase	
U89648		<i>Corynebacterium glutamicum</i> unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the <i>Brevibacterium lactofermentum</i> tryptophan operon," <i>Nucleic Acids Res.</i> , 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of <i>Corynebacterium glutamicum</i> and possible mechanisms for modulation of its expression," <i>Mol. Gen. Genet.</i> , 212(1):112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of <i>Corynebacterium glutamicum</i> : Molecular cloning, nucleotide sequence, and expression," <i>Mol. Gen. Genet.</i> , 218(2):330-339 (1989); Lepiniec, I. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," <i>Plant Mol. Biol.</i> , 21 (3):487-502 (1993)
X17113	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine structural analysis of the <i>Corynebacterium glutamicum</i> fda gene: structural comparison of <i>C. glutamicum</i> fructose-1, 6-bisphosphate aldolase to class I and class II aldolases," <i>Mol. Microbiol.</i>
X53993	dapA	L-2,3-dihydrodipicolinate synthetase (EC 4.2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from <i>Corynebacterium glutamicum</i> ," <i>Nucleic Acids Res.</i> , 18(21):6421 (1990)

GenBank <sup>TM</sup> Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Ciancioffo, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diphtheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of <i>lambdacorynebphage</i> ," <i>FEMS Microbiol. Lett.</i> , 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the <i>Corynebacterium glutamicum</i> lysA gene," <i>Mol Microbiol.</i> , 4(11):1819-1830 (1990)
X55904	trpL; trpE	Putative leader peptidase; anthranilate synthase component I	Heery, D.M. et al. "Nucleotide sequence of the <i>Corynebacterium glutamicum</i> trpE gene," <i>Nucleic Acids Res.</i> , 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the <i>Corynebacterium glutamicum</i> threonine synthase gene," <i>Mol Microbiol.</i> , 4(10):1693-1702 (1990)
X56075	attB-related site	Attachment site	Ciancioffo, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diphtheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of <i>lambdacorynebphage</i> ," <i>FEMS Microbiol. Lett.</i> , 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit, Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from <i>Corynebacterium glutamicum</i> ," <i>Mol Microbiol.</i> , 5(5):1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC-alpha and lysC-beta overlap and are adjacent to the aspartate beta-semialdehyde dehydrogenase gene asd in <i>Corynebacterium glutamicum</i> ," <i>Mol Gen Genet.</i> , 224(3):317-324 (1990)
X59403	gap-pgk; tpi	Glyceraldehyde-3-phosphate: phosphoglycerate kinase, triosephosphate isomerase	Eikmanns, B.J. "Identification, sequence analysis, and expression of a <i>Corynebacterium glutamicum</i> gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomerase," <i>J Bacteriol.</i> , 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E.R. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> gdh gene encoding glutamate dehydrogenase," <i>Mol Microbiol.</i> , 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Scep-Feldhaus, A.H. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> lysI gene involved in lysine uptake," <i>Mol Microbiol.</i> , 5(12):2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	copI	Psl protein	Joliff, G. et al. "Cloning and nucleotide sequence of the cspI gene encoding PSI, one of the two major secreted proteins of <i>Corynebacterium glutamicum</i> : The deduced N-terminal region of PSI is similar to the Mycobacterium antigen 85 complex." <i>Mol. Microbiol.</i> , 6(16):2349-2362 (1992)
X66112	gli	Citrate synthase	Eikmanis, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the <i>Corynebacterium glutamicum</i> gltA gene encoding citrate synthase." <i>Microbiol.</i> , 140:1817-1828 (1994)
X67737	denB	Dihydropyridine reductase	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in <i>Corynebacterium glutamicum</i> ." <i>Mol. Microbiol.</i> , 9(1):97-109 (1993)
X69103	csp2	Surface layer protein PS2	Bonamy, C. et al. "Identification of IS1206, a <i>Corynebacterium glutamicum</i> IS3-related insertion sequence and phylogenetic analysis." <i>Mol. Microbiol.</i> , 14(3):571-581 (1994)
X69104		IS3 related insertion element	Patek, M. et al. "Leucine synthesis in <i>Corynebacterium glutamicum</i> : enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis." <i>Appl. Environ. Microbiol.</i> , 60(1):133-140 (1994)
X70959	leuA	Isopropylmalate synthase	Eikmanns, B.J. et al. "Cloning sequence analysis, expression, and inactivation of the <i>Corynebacterium glutamicum</i> icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme." <i>J. Bacteriol.</i> , 177(3):774-782 (1995)
X71489	icd	Isocitrate dehydrogenase (NADP+)	
X72855	GDHA	Glutamate dehydrogenase (NADP+)	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of <i>Corynebacterium glutamicum</i> encoding resistance to 5-methyltryptophan." <i>Biochem. Biophys. Res. Commun.</i> , 201(3):1255-1262 (1994)
X75083, X70584	mitA	5-methyltryptophan resistance	Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of <i>Corynebacterium glutamicum</i> and <i>Brevibacterium lactofermentum</i> ." <i>Appl. Microbiol. Biotechnol.</i> , 42(4):575-580 (1994)
X75085	iccA		Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from <i>Corynebacterium glutamicum</i> and biochemical analysis of the enzyme." <i>J. Bacteriol.</i> , 176(12):3474-3483 (1994)
X75504	aceA; thiX	Partial isocitrate lyase?	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes." <i>Antonie Van Leeuwenhoek</i> , 64:285-305 (1993)
X76875		ATPase beta-subunit	



GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and A TP-synthase beta-subunit genes," <i>Antonie Van Leeuwenhoek</i> , 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from <i>Corynebacterium glutamicum</i> ," <i>DNA Seq.</i> 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase: sequence analysis," <i>Microbiology</i> , 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera <i>Rhodococcus</i> and <i>Norcardia</i> and evidence for the evolutionary origin of the genus <i>Norcardia</i> from within the radiation of <i>Rhodococcus</i> species," <i>Microbiol.</i> , 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronmeyer, W. et al. "Structure of the gluABCD cluster encoding the glutamate uptake system of <i>Corynebacterium glutamicum</i> ," <i>J. Bacteriol.</i> , 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehmann, A. et al. "Analysis of different DNA fragments of <i>Corynebacterium glutamicum</i> complementing dapE of <i>Escherichia coli</i> ," <i>Microbiology</i> , 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R. et al. "Phylogeny of the genus <i>Corynebacterium</i> deduced from analyses of small-subunit ribosomal DNA sequences," <i>Int. J. Syst. Bacteriol.</i> , 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, J. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," <i>J. Bacteriol.</i> , 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, J. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," <i>J. Bacteriol.</i> , 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus <i>Corynebacterium</i> based on 16S rRNA gene sequences," <i>Int. J. Syst. Bacteriol.</i> , 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehmann, A. et al. "Functional analysis of sequences adjacent to dapE of <i>Corynebacterium glutamicum</i> reveals the presence of aroP, which encodes the aromatic amino acid transporter," <i>J. Bacteriol.</i> , 177(20):5991-5993 (1995)

GenBank <sup>TM</sup> Accession No.	Gene Name	Gene Function	Reference
X86157	argB, argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in <i>Corynebacterium glutamicum</i> : enzyme evolution in the early steps of the arginine pathway," <i>Microbiology</i> , 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase, acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase and acetate kinase," <i>Microbiology</i> , 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting <i>Arthrobacter aureus</i> C70," <i>J. Bacteriol.</i> , 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R. M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of <i>Corynebacterium glutamicum</i> ," <i>J Biol Chem</i> , 271(10):5398-5403 (1996)
X93514	belP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the <i>Corynebacterium glutamicum</i> belP gene, encoding the transport system for the compatible solute glycine betaine," <i>J Bacteriol</i> , 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of <i>Corynebacterium glutamicum</i> , encoding two enzymes involved in L-lysine synthesis," <i>Biotechnol Lett</i> , 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein, Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from <i>Corynebacterium glutamicum</i> ," <i>Mol Microbiol</i> , 22(5):815-826 (1996)



GenBank <sup>TM</sup> Accession No.	Gene Name	Gene Function	Reference
X96580	panB, panC, xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase, pantoate-beta-alanine ligase; xylulokinase	Sahn, H. et al. "D-pantothenate synthetase in <i>Corynebacterium glutamicum</i> and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," <i>Appl Environ Microbiol</i> , 65(5):1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer <i>Brevibacterium lactofermentum</i> ( <i>Corynebacterium glutamicum</i> ATCC 13869)," <i>Gene</i> , 198:217-222 (1997)
X99289		Elongation factor P	
Y00140	thrB	Homoserine kinase	Malcos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the <i>Brevibacterium lactofermentum</i> ," <i>Nucleic Acids Res</i> , 15(9):3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from <i>Corynebacterium glutamicum</i> ," <i>Nucleic Acids Res</i> , 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the <i>Brevibacterium lactofermentum</i> ," <i>Nucleic Acids Res</i> , 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the <i>Corynebacterium glutamicum</i> hom-thrB operon," <i>Mol Microbiol</i> , 2(1):63-72 (1988)
Y08964	murC, fisQ/divD; fisZ	UDP-N-acetylmuramate-alanine ligase; division initiation protein or cell division protein; cell division protein	Honrubia, M.P. et al. "Identification, characterization, and chromosomal organization of the fisZ gene from <i>Brevibacterium lactofermentum</i> ," <i>Mol Gen Genet</i> , 259(1):97-104 (1998)
Y09163	puuP	High affinity proline transport system	Peter, H. et al. "Isolation of the puuP gene of <i>Corynebacterium glutamicum</i> proline and characterization of a low-affinity uptake system for compatible solutes," <i>Arch Microbiol</i> , 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from <i>Corynebacterium glutamicum</i> : characterization, expression and inactivation of the pyc gene," <i>Microbiology</i> , 144:915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patel, M. et al. "Analysis of the leuB gene from <i>Corynebacterium glutamicum</i> ," <i>Appl Microbiol Biotechnol</i> , 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynebacteriophage Phi-16: The construction of an integration vector," <i>Microbiol</i> , 145:539-548 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP." <i>J. Bacteriol.</i> , 180(22):6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I." <i>FEMS Microbiol Lett.</i> , 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	Moreau, S. et al. "Analysis of the integration functions of &phi;304L. An integrase module among corynephages." <i>Virology</i> , 255(1):150-159 (1999)
Y18059		Attachment site Corynephage 304L	Oguiza, J A. et al. "A gene encoding arginyl-(tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum. Regulation of argS-lysA cluster expression by arginine." <i>J Bacteriol.</i> , 175(22):7356-7362 (1993)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diamino pimelate decarboxylase (partial)	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function." <i>J. Bacteriol.</i> , 175(9):2743-2749 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase." <i>Appl Environ Microbiol.</i> , 60(7):2209-2219 (1994)
Z29563	thrC	Threonine synthase	Oguiza, J A. et al. "Multiple sigma factor genes in Brevibacterium lactofermentum. Characterization of sigA and sigB." <i>J Bacteriol.</i> , 178(2):550-553 (1996)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	Oguiza, J A. et al. "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dmdK gene." <i>Gene</i> , 177:103-107 (1996)
Z49822	sigA	SigA sigma factor	Oguiza, J A. et al. "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB." <i>J. Bacteriol.</i> , 178(2):550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diaphtheria toxin regulatory protein	Concia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869." <i>Gene</i> , 170(1):91-94 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	
Z66534		Transposase	

A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: *Corynebacterium* and *Brevibacterium* Strains Which May be Used in the Practice of the Invention

Strain	Genus	Species	ATCC	From	Num	Sub	NCIM	CIS	NCIC	DSMZ
Brevibacterium	ammonia	ammonia	21054							
Brevibacterium	ammonia	ammonia	19350							
Brevibacterium	ammonia	ammonia	19351							
Brevibacterium	ammonia	ammonia	19352							
Brevibacterium	ammonia	ammonia	19353							
Brevibacterium	ammonia	ammonia	19354							
Brevibacterium	ammonia	ammonia	19355							
Brevibacterium	ammonia	ammonia	19356							
Brevibacterium	ammonia	ammonia	21055							
Brevibacterium	ammonia	ammonia	21077							
Brevibacterium	ammonia	ammonia	21553							
Brevibacterium	ammonia	ammonia	21580							
Brevibacterium	ammonia	ammonia	39101							
Brevibacterium	ammonia	ammonia	21196							
Brevibacterium	ammonia	ammonia	21792	P928						
Brevibacterium	ammonia	ammonia	21474							
Brevibacterium	ammonia	ammonia	21129							
Brevibacterium	ammonia	ammonia	21518							
Brevibacterium	ammonia	ammonia			B11474					
Brevibacterium	ammonia	ammonia			B11472					
Brevibacterium	ammonia	ammonia	21127							
Brevibacterium	ammonia	ammonia	21128							
Brevibacterium	ammonia	ammonia	21427							
Brevibacterium	ammonia	ammonia	21475							
Brevibacterium	ammonia	ammonia	21517							
Brevibacterium	ammonia	ammonia	21528							
Brevibacterium	ammonia	ammonia	21529							
Brevibacterium	ammonia	ammonia			B11477					

Brevibacterium	flavum			B11478					
Brevibacterium	flavum	21127							
Brevibacterium	flavum			B11474					
Brevibacterium	halali	15527							
Brevibacterium	ketoglutaricum	21004							
Brevibacterium	ketoglutaricum	21089							
Brevibacterium	ketosoreductum	21914							
Brevibacterium	lactofermentum					70			
Brevibacterium	lactofermentum					74			
Brevibacterium	lactofermentum					77			
Brevibacterium	lactofermentum	21798							
Brevibacterium	lactofermentum	21799							
Brevibacterium	lactofermentum	21800							
Brevibacterium	lactofermentum	21801							
Brevibacterium	lactofermentum			B11470					
Brevibacterium	lactofermentum			B11471					
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	21420							
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	31269							
Brevibacterium	linens	9174							
Brevibacterium	linens	19391							
Brevibacterium	linens	8377							
Brevibacterium	paraffinolyticum					11160			
Brevibacterium	spec.						717.73		
Brevibacterium	spec.						717.73		
Brevibacterium	spec.	14604							
Brevibacterium	spec.	21860							
Brevibacterium	spec.	21864							
Brevibacterium	spec.	21865							
Brevibacterium	spec.	21866							
Brevibacterium	spec.	19240							









ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Colección Española de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelmcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H. et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4<sup>th</sup> edn), World Federation for culture collections world data center on microorganisms, Saitama, Japan.

>>RXA00459-amino acid sequence  
(1-987, translated) 329 residues

VCTR AAGGGAVTLK RARELRKKRG RMAARIADSV MAGELLHATG AIDRELNAV T RDSDRVVIAA  
VRRSWATGFS RALMAMAASL GTVSIVISGH LEVSEVAGIM MLLGVLATPV AELGRVVEYR QNYKAATRIL  
IPLLQRGSEF KHSQQKLPGL QATEGIPGVY VKGISALPGE RIYLHGSADA TRKWVTSLSA MEEGTDVIVN  
GQRLSQLPLK QRRALIGIAS AHHHLSRGSV SRLVGLRVPD ATVEEIEQAL EQVGLNNTGK QRLKNGGHPW  
STSQINKLKI ASATLRTPPL LVLEGITPEN LLNYPGVIIIS TVQENPSETW RQVNI

>RXA00459-nucleotide sequence A: upstream

AGGCGTGGCTGTTACTGTGCCACTGCTCAT

>RXA00459-nucleotide sequence B: coding region

GTGTGTACCCGTGCTGCCGGTGGTGGCGCGGTGACTTTGAAAAGAGCACGTGAACTACGCAAAAAACGTGGACGCAT  
GGCTGCGCGGATCGCAGATTCTGTCATGGCTGGAGAATTACTGCACGCAACAGGAGCAATAGACCGTGAGCTCAATG  
CAGTCACCCGAGATTCCGACCGAGTGGTGATAGCTGCTGTAAGACGTTCCCTGGGCCACCGGTTTTAGCCGCGCATTG  
ATGGCCATGGCAGCCTCGCTTGGCACTGTCAGCATTGTGATTTCTGGCCACCTGGAAGTAAGTGAGGTTGCGGGAAT  
AATGATGCTTCTTGGCGTTCTTGCCACTCCAGTTGCAGAACTTGGCCGCGTGGTGGAATATCGCCAAAATTATAAAG  
CCGCGACACGCATCCTGATTCCACTTCTGCAACGAGGCTCAGAATTTAAACACTCCCAACAAAAACTACCCGGGTTG  
CAAGCAACAGAAGGAATCCCCGGTGTCTATGTCAAAGGTATTTCCGCCCTTCTGGAGAACGGATCTACCTCCACGG  
CTCTGCAGATGCGACGAGAAAATGGGTACCTCGTTGTCTGCAATGGAGGAAGGCACAGATGTAATAGTCAACGGTC  
AAAGGCTTTTCGAGCTTCCTTTGAAACAACGACGGGCCCTCATCGGAATCGCCTCAGCACACCACCACTTAAGCCGT  
GGTTCAGTATCGCGCTGGTTGGTTTGCAGTGCCGGATGCCACCGTGAAGAAATTGAGCAAGCACTGGAACAAGT  
TGCTCTGAACAACACCGGGAACAACGCTTGAAAAACGGCGGACACCCCTGGAGTACTTCGAGATCAACAACTGA  
AAATTGCCAGCGCCACCCTTCGAACCCACCGCTTTTGGTACTTGAAGGCATCACCCCTGAAAACCTCCTCAACTAT  
CCCGAGTGATCATCTCCACCGTTCAGGAGAACCCTCCGAAACATGGCGGCAAGTGAACATC

>RXA00459-nucleotide sequence C: downstream

TAATCTAGAAACATGGCAGGACG

>>RXA00410-amino acid sequence  
(1-666, translated) 222 residues

MMIYGKGS TE VRALDGISVQ IQSDKWTSIM GQSGSGKTTL LQCLSGLAQP TSGRVTLNKN NITLSSLSEN  
KRAKLRRTHI SMVFQDFNLV PILSVKDNIL LPLRLAHRV DKQWFEHITS VLKIDNMRH LPGELSGGQQ  
QRAAIARALM SRPDIVIADE PTGSLDSVTS DAVLNLFRSI VDDFGQSLVF VTHDKDAAHR GDVLITMRDG  
KIIDTADLRV GR

>RXA00410-nucleotide sequence A: upstream

GTGTTGATGCGTTAGTCCACCCACGCAGCTACGCCCCAAAGGAATAATCTTGAACCCTGCCACAGATAACGCTCCGC  
CGGTCCTTTTCAGCCCCAAGATCTC

>RXA00410-nucleotide sequence B: coding region

ATGATGATCTATGGAAAAGGATCAACAGAAGTTCGGGCTCTCGATGGCATTCTGTACAGATTCAGTCCGACAAATG  
GACCTCCATCATGGGGCAATCAGGCTCTGGCAAACAACCTCTGTTGCAGTGCCTTTCCGGATTGGCGCAGCCAACCT  
CAGGCAGAGTGACACTGAACAAAAACAACATCACGTTGAGCTCCCTGTCAGAAAATAAGCGTGCCAAGCTGCGTCGC  
ACGCACATCAGCATGGTGTTCAGGATTTCAACTTGGTGCCTATTTTGTTCGGTGAAGGACAATATTTTGTCTGCCGTT  
GCGTCTTGCGCATCGCAGGGTGGATAAGCAGTGGTTTGAACACATCACCAGTGTGTTGAAGATTGATAATCGTATGC  
GCCATTTGCCTGGGGAGCTGTCTGGCGGTCAGCAACAACGCGCCGCGATTGCCCCGGGCGTTGATGTCTAGGCCCCGAT  
ATTGTCATTGCGGATGAGCCAACAGGAAGTTTGGATTCCGTCACCAGCGATGCAGTGTGAATTTGTTCCGCAGCAT  
TGTTGATGATTTTGGGCAGTCACTTGTGTTTGTACCCACGATAAAGATGCTGCTCACCGTGGTGACGTGTTGATCA  
CAATGCGTGATGGCAAGATCATCGATACGGCAGATTGCGGGTGGGGCGT

>RXA00410-nucleotide sequence C: downstream

TAATGTTTCAGGCTTGCTTTTCGCT

>>RXA00526-amino acid sequence  
(1-690, translated) 230 residues

MSLIEMRNIV KTYNIGSEGE LTVLHGVDHF VDRGEFVSVV GTSGSGKSTM MNIIGLLDKP TDGTYTLDGV  
DVLDISDDAL ASHRAKSIGF VFQNFNLIGR IDALKNVEMP MMYAGIPAKQ RRSRAVELLE MVGMGERLNH  
EPNELSGGQK QRVAIARALA NDPEIILADE PTGALDSATG RMVMDIFHQL NKEQGKTIVE ITHNPELADE  
SDRVVTMVDG RIIGSEVKHS

>RXA00526-nucleotide sequence A: upstream

GGTGGAGCAGGCGGCGCTCCTTTTAGTCCTGCGGCCCTTTTGACCCTGCAGCCCCTGCCGTTTCTGCCAAGCAAA  
CCGTGGGCCAGGTGATTTAGCCT

>RXA00526-nucleotide sequence B: coding region

ATGAGCCTCATCGAAATGCGAAATATTGTCAAGACCTACAACATTGGATCTGAAGGTGAACTCACCGTGTTGCACGG  
TGTGGATTTCCATGTGGACCGTGGCGAATTCGTGTGCGGTTGTGGGTACGTCCGGCTCAGGTAAATCAACGATGATGA  
ACATCATTGGGTTGTTGGATAAGCCAACTGATGGCACGTACACCTTGGATGGCGTGATGTGTTGGATATCAGCGAT  
GATGCTTTGGCGAGCCACCGCGCTAAATCGATTGGTTTTGTGTTTCAGAACTTCAATCTGATTGGCCGGATCGATGC  
GTTGAAGAATGTGGAAATGCCCATGATGTATGCGGGCATTCCGGCTAAGCAGCGGAGAAGTCGTGCGGTTGAATTAT  
TGGAAATGGTCCGGATGGGTGAGCGTCTCAACCATGAGCCCAATGAGCTTTCGGGTGGTCAGAAGCAGCGCGTGGCC  
ATTGCTCGCGCGTTGGCGAACGATCCTGAGATCATTCTTGCTGATGAACCAACTGGTGGCTTGGATTCTGCAACGGG  
CCGGATGGTGATGGATATTTTCCACCAGCTCAACAAGGAGCAGGGCAAAACCATCGTGTTTATTACTCACAACCCCTG  
AGCTTGCTGATGAATCTGATCGGGTGGTCACCATGGTTGACGGGCGCATCATTGGGTCTGAGGTGAAACACTCA

>RXA00526-nucleotide sequence C: downstream

TGAGCCTTGCGAATCAATTCTT

>>RXA00735-amino acid sequence  
(1-669, translated) 223 residues

GPTGAGKTTL VNLIMRFYDI NSGSITLGET AQDAVDIRTM AREDLRSRTG MVLQDTWLFA GTIRDNILYG  
RPEATEEEML AASKAAYVDR FVRSLEPGYD TVLDDEAMNL SVGERQLITI ARAFLANPRL LILDEATSSV  
DTRTELLIQR AMSKLRQDRT AFVIAHRLST IRDANLIIMM KDGEIVEQGN HRELMALEGA YWELYN SQFN  
APAKEELQAD GDH

>RXA00735-nucleotide sequence B: coding region

GGTCCCACCGGTGCGGGCAAGACCACATTGGTGAATCTGATCATGCGTTTCTACGACATCAACAGCGGTTCCATCAC  
TCTTGGTGAAACAGCACAAAGACGCCGTGGATATCCGCACCATGGCTAGAGAAGATCTGCGATCACGAACCGGCATGG  
TGTTGCAGGATACGTGGCTGTTTGCCGGAACCATCAGGGATAACATTCTTTACGGTAGACCTGAAGCAACTGAGGAA  
GAAATGCTTGCTGCGTCCAAGGCCGCCTACGTGGATCGTTTTGTCCGTTCCCTGCCAGAAGGCTACGACACCGTACT  
TGATGATGAAGCCATGAACCTATCGGTGGGTGAACGCCAGCTGATCACCATCGCGCGTGCATTCTTGGCTAATCCCC  
GACTGCTGATTCTGGATGAAGCCACCTCATCGGTGGATACCGGTACCGAATTGTTGATTCAGCGCGCCATGTCCAAG  
CTGCGCCAAGACCGCACCGCCTTCGTCATCGCGCACCGGTTGTCCACGATTCTGTGATGCCAACCTGATTTTGATGAT  
GAAAGACGGCGAGATCGTGGAGCAGGGCAATCACCGTGAGTTGATGGCCCTGGAGGGCGCATATTGGGAGTTGTATA  
ACTCCCAATTCAACGCCCCCGCGAAAGAAGAATTACAGGCTGACGGAGATCAC

>RXA00735-nucleotide sequence C: downstream

TGATGATTTCTTCTTAGGCTTTC



>>RXA00734-amino acid sequence  
(1-453, translated) 151 residues

RHLRYGNEDA TETQLWQALA IAQAADFVRE MPEGLDSEIA QGGTNVSGGQ RQLAIARAL LKQPEIYIFD  
DSFSALDVST DAALRRALST NLPDATKLIV AQRVSTIRDA DQIVVLDNGE VVGIGTHTNL LNTCGTYREI  
VESQETAQAQ S

>RXA00734-nucleotide sequence B: coding region

AGGCACCTGCGTTATGGCAATGAAGATGCCACGGAAACGCAGCTGTGGCAGGCGCTTGCAATTGCTCAGGCGGCGGA  
CTTTGTGCGTGAGATGCCAGAGGGTCTTGATTCTGAGATTGCTCAGGGTGGAACCAATGTTTCTGGTGGTCAGCGCC  
AGCGACTAGCCATTGCCAGGGCGTTGTTGAAGCAACCTGAGATCTATATTTTCGACGATTCTTCTCCGCCCTCGAT  
GTGAGCACAGACGCCGCTCTTCGCCGAGCGCTGAGCACCAACCTGCCGGATGCAACCAAGTTGATTGTCGCCCAGCG  
TGTCAGCACGATTGAGATGCCGATCAGATTGTGGTGCTTGATAACGGCGAGGTTGTCGGTATTGGAACGCACACGA  
ATTTGCTGAACACGTGCGGTACCTACCGTGAAATTGTTGAATCCCAAGAGACTGCGCAGGCGCAATCA

>RXA00734-nucleotide sequence C: downstream

TGAGTAATACTGCAGGCCCCCGC

>>RXA00733-amino acid sequence  
(1-408, translated) 136 residues

MSNTAGPRGR SHQADAAPNQ KAQNFGPSAK RLFILGHDR NTLIFVIFLA VLSVGLTVLG PWLLGKATNV  
VFEGFLSKRM PAGASKEDII AQLQAAGKHN QASMEDMNL VPGSGIDFEK LAMILGLVIG AYLIRS

>RXA00733-nucleotide sequence A: upstream

ACGGCGAGGTTGTCGGTATTGGAACGCACACGAATTTGCTGAACACGTGCGGTACCTACCGTGAAATTGTTGAATCC  
CAAGAGACTGCGCAGGCGCAATC

>RXA00733-nucleotide sequence B: coding region

ATGAGTAATACTGCAGGCCCCCGCGGGCGTTCCCATCAGGCAGACGCCGCGCCGAATCAAAAAGGCACAGAATTTTCGG  
ACCATCTGCCAAAAGGCTTTTCGGAATTCTAGGCCATGACCGTAACACCTTAATTTTGTATCTTCCTAGCCGTCC  
TGAGCGTTGGACTTACCGTCTTGGGCCCCATGGTTGCTGGGTAAAGCCACCAACGTGGTGTGTTGAAGGATTCCTATCT  
AAGCGCATGCCGGCTGGTGCGTCAAAGGAAGATATCATCGCGCAGTTGCAGGCTGCAGGTAAACATAATCAGGCTTC  
CATGATGGAAGACATGAACCTTGTTCCAGGCTCAGGCATTGATTTTGAAAAATTAGCCATGATCCTCGGACTGGTGA  
TCGGTGCTTATCTCATTCTAGC

>>RXA00878-amino acid sequence  
(1-1863, translated) 621 residues

MRLLGRI LKT TSALWPYYLG IIVVSIVIAA LSLSPFILR EATDSIVSAV TGSNTVDAVT RTIIFLALAL  
FVASFLNTVM TNIGGYIGDV MASRMROI LA TRYAKLLAL PQKYFDNQVT GTIIARLDRS INGITOQFMS  
FSNFFPMLI TMVAVLIISA IFYWPLAILL AMLFPIYMWL TALTSKRWQK YEGEKNHEID VANGRFAEVV  
GQVKVVKSFV AETRELADFG GRYGKTVAIT RPQSGWWHRM DTLRGAALNI IFLAIHLLIF YRTLHGHFTI  
GDMVMLIQLV TMAQQPVYMM SYIVDSAQRA IAGSRDYFEV MAQQVEPTAN KELVDATLAS DTPRISVGTP  
AALPAGEPAM EFKNVTFAYE EGKPVISDVS ITARHGERIA LVGESGGGKS TLVNLGLLY KPNSSGLAVC  
GVDVKDLTSE ELRASVGVVF QDASLFSGSI AENIAYGRPG ATREEIIEVA KKANAEHIFIS AFPEGYETVV  
GERGLKLSGG QKQRVSVARA MLKDAPLLVL DEATSALDTK SEQAVQAGLE QLMENRTTLM IAHRLSTIAG  
VDTIVTIQNG RVEEVGSPTE LAVSGGIYSE LLRLTNSTAE ADRERLRAFG FTGDAPAEED D

>RXA00878-nucleotide sequence A: upstream

CGAGATTAGGTCCGCTTCAGTTGTGGTGGCTCCGAATCTGATGAACAATGATCATTCTTAATTCATTACATCTTTA  
TCAAAGAGAGCCACCACCTACTA

>RXA00878-nucleotide sequence B: coding region

ATGCGACTTCTTGGTCGAATTTTAAAAACCACGTCTGCGCTTTGGCCCTACTATCTCGGAATTATCGTCGTATCCAT  
TGTGATCGCGCGCTTGTCTGCTGCTGTCGCCGTTTATTCTCCGCGAAGCAACAGATTCCATTGTTTCTGCAGTAACCG  
GATCTAACACCGTCGATGCAGTTACTCGCACTATTATTTCTTAGCTTTAGCCCTGTTTGTGCGCAAGCTTCCTCAAT  
ACGGTGATGACCAACATCGGTGGCTACATCGGTGATGTCTGATGGCAGCTCGATGCGCCAGATTCTGGCCACGCGCTA  
TTACGCAAAGCTGTTGGCGCTGCCTCAGAAGTATTTTGATAATCAGGTACCGGCACCATCATCGCCCGCTTGATC  
GATCAATCAACGGCATCACGCAGTTTCATGCAGAGCTTCTCCAACAACCTTCTTCCCATGCTCATCACCATGGTGGCA  
GTGCTGATTATTTCCGCGATTTTCTACTGGCCTCTGGCAATTCTGCTGGCCATGTTGTTCCCGATTACATGTGGCT  
GACGGCGTTGACATCGAAACGCTGGCAGAAATATGAGGGCGAGAAAAACCATGAAATCGACGTGGCTAACGGCCGCT  
TCGCTGAGGTTGTGCGCCAGGTCAAGGTTGTTAAATCATTTCGTCGCAGAGACCCGCGAGCTGGCTGATTTCCGTGGG  
CGTTACGGCAAAACAGTAGCGATTACCGCGCCGAATCCGGTTGGTGGCACCAGCATGGATACTCTCCGTGGCGCGGC  
ACTAAATATCATCTTCTGCGCATTCACCTGCTGATTTTCTACCGCACCTTGACAGGCCATTTACCATCGGCGACA  
TGGTCATGCTCATCCAGCTTGTCAACATGGCGCAGCAACCGGTGTACATGATGAGCTACATCGTCGACTCCGCGCAG  
CGCGCCATCGCCGCTCCCGCGACTACTTCGAGGTTCATGGCGCAGCAGGTTCGAGCCCACCGCCAATAAGGAGCTTGT  
CGACGCCACCCCTCGCCTCAGACACTCCACGCATCAGTGTGGGCACGCCCCGCGCTGCCGCTGGAGAACCAGCGA  
TGGAATTCAAAAACGTCACCTTCGCCTACGAAGAAGGCAAGCCGGTTATTTCCGACGTGTCCATTACCGCCCGCCAC  
GGCGAGCGCATCGCGTTGGTTCGGTGAATCCGGCGCGGTAAATCCACCCTGGTCAACCTTCTGTTAGGTCTGTACAA  
ACCAAACAGCGGCAGCCTTGCAATGTGGCGTGGATGTTAAAGATCTGACTTCCGAGGAACCTTCGCGCATCCGTGG  
GTGTGGTCTTCCAGGACGCCAGCTTGTCTCTGGATCTATTGCAGAAAACATCGCCTACGGTCGCCCGAGGTGCCACC  
CGCGAAGAGATCATCGAAGTGGCTAAGAAAGCCAACGCACATGAGTTCATTTCGCGCTTCCCTGAAGGATATGAAAC  
CGTCGTCGGTGAACGCGGACTCAAACCTTCTGGTGGCCAGAAGCAGCGCTCTCTGTGGCACGGGCCATGCTTAAAG  
ATGCCCCACTTCTTGTCTCGATGAAGCCACCTCTGCACTGGATACCAAGTCTGAGCAGGCAGTCCAAGCCGGTTT  
GAACAGCTGATGGAACCGCACCACCTTAATGATCGCCACCGCCTGTCCACCATCGCAGGCGTCGATACCATCGT  
GACCATCCAAAACGGACGGGTTGAAGAGGTTCGATCTCTACCGAGCTCGCAGTCTCAGGCGGTATCTATTCCGAAC  
TGCTGCGCCTGACCAACTCCACAGCAGAAGCCGACGGGAGCGTCTGCGCGCCTTGGTTTCACTGGCGATGCACCA  
GCTGAAGAAGAGGAC

>RXA00878-nucleotide sequence C: downstream

TAGCCCCGCGAAAGAACAATCCC

>>RXA01191-amino acid sequence  
(1-1407, translated) 469 residues

VSLDANTIET AGRGDVISRI ADDSREVSTA ASTVVPLMVQ AGFTTVISAF GMAAVDWRLG LVGLVAIPLY  
WTTLRVYLPR SGPLYTRERE AFGVRTQRLV GAVEGAETLR AFRAEDTELK RIDAASGEAR DISISVFRFL  
TWAFSRNNRA ECITLVILIG TGFYLVNIDL VTVGAVSTAA LIFHRLFGPI GTLVGMFSDI QSASASLIRM  
VGVINAASNQ VSGTSPASAS TALTLFDVSH HYHTAPVIKN ASVQLPEGEH IAIVGATGAG KSTLALIAAG  
LLSPTSGQVA LGGSSFSNVE PEALRQKIAM VSQEIHCFRG SVLDNLRIAR PEATDADIHA VLADIGDSWL  
ERLPQGIDTI VGDGAFLRLTS VENQIMALAR VHLADLAIVI LDEATAESGS DHAKQLEDAA LKV TENRSAI  
IVAHRLNQAK TADRIIVMDS GEIIESGTHE ELRAIGGRYE QLWTAWSAR

>RXA01191-nucleotide sequence B: coding region

GTGAGTTTGGATGCGAACACGATTGAAACGGCGGGCGCGCGGCGATTTTCGCGTATCGCGGATGATTTCGCGGGA  
GGTGTCCACTGCGGCGAGCACCGTGGTGGCGCTGATGGTGCAGGCGGGCTTTACCGTGGTGATTTCGCGGTTTGGCA  
TGGCGGCGGTTGATTGGCGCCTCGGCCTTGTTCGGTTTGGTTCGCGATCCCGCTGTATTGGACCACGTTGCGCGTCTAT  
TTACCCCGCTCAGGTCCGCTTTTATACGCGTGAGCGCGAGGCCTTTGGGGTGCGCACGCAGCGGCTTGTTCGGCGCAGT  
CGAAGGCGCGGAAACCTTTCGCGGCTTTTCGCGCAGAGAATACAGAATTAAAGCGTATCGACGCAGCCTCCGGCGAAG  
CCCGCGACATTTCCATTTCTGTTTTTCAGGTTCTCATATGGGCATTTTCCCGCAACAACCGCGCGGAATGCATCACC  
CTCGTGCTCATCTTGGGCACCGGCTTTTACCTGGTCAACATCGATCTGGTCAACCGTCCGGCGCAGTCTCAACCGCCGC  
ACTGATCTTCCACCGACTCTTTCGGTCCAATCGGCACGCTCGTGGGCATGTTCTCCGACATCCAATCCGCCAGCGCAT  
CGCTGATCCGCATGGTGGGCGTTATTAACGCGGCATCGAACCAGGTGAGCGGCACCTCGCCGGCGTCTGCCAGCACC  
GCTTTAACGCTTTTCGACGTCTCCCACTATCACACTGCACCCGTCATCAAGAATGCATCCGTGCAGCTGGAACC  
AGGGGAACACATCGCCATTGTGGGTGCGACCGGCGCTGGTAAAGCAGCGTCCGCCCTCATTGCGGCAGGCCTGCTCA  
GCCCCAATTCGCGGCAGGTGGCTCTCGGCGGATCGAGTTTTTCTAACGTGGAACCGGAAGCATTGCGCCAGAAGATC  
GCGATGGTCAGCCAAGAAATCCACTGCTTCCGAGGATCTGTTTTAGATAATCTTCGTATCGCACGCCCCGAAGCCAC  
CGATGCGGACATCCACGCCGTTCTCGCCGATATTGGTGATTCTGGTTGGAGCGCTTACCGCAAGGCATAGACACCA  
TCGTGGGTGATGGCGCTTTCCGTTTAACTCTGTGGAACACAGATCATGGCGCTTGCTCGCGTACATTTGGCCGAC  
CTAGCAATCGTCATCCTTGATGAAGCAACGGCTGAATCAGGCTCTGATCATGAAAACAGCTTGAAGATGCAGCCCT  
TAAAGTCACTGAAAACAGATCAGCCATCATCGTGGCTCACCGCCTCAACCAAGCGAAAACCGCCGATCGCATCATCG  
TCATGGACTCCGGAGAAATCATAGAATCTGGAACCCATGAAGAGCTTCGAGCGATCGGCGGTTCGATATGAACAACG  
TGGACTGCGTGGTCTGCGCGC

>RXA01191-nucleotide sequence C: downstream

TAATTAGCCACCCAAGACCACGC

>>RXA01212-amino acid sequence  
(1-813, translated) 271 residues

GLNFHVQRGE VFGLLGTTGA GKTSTLEVIE GLSAPSSGTV RISGLDFVAD RAILRPELGI MLQSGGLPSQ  
LTVAETMDMW HGTCTYPRAI KDVLADVDLL HRENVKVGAL SGGEQRRLDL ACALLGDPSI LFLDEPTTGL  
DPESRRHTWQ LLLDLKQRGV TMLTTHYLE EAEFLCDRIA IMNAGEIAVE GTLDELVARE KSIISFVLRG  
GQVELPVLSC AEIIRDNNHV RIATTTLQOH TLEILTAAE TGIALEGFAA KPATLESVFM D

>RXA01212-nucleotide sequence B: coding region

GGCCTGAATTTCCATGTACAGCGCGGTGAAGTATTTGGTCTGCTCGGCACCAACGGGGCCGGCAAAACCTCCACCTT  
GGAAGTCATCGAAGGACTTTCCGCACCCAGCTCCGGCACCGTGCGCATCTCCGGGCTTGACCCCGTTGCCGACCGCG  
CGATCCTGCGCCCCGAGCTCGGCATCATGCTGCAATCAGGCGGCCTGCCATCACAGCTCACCGTCGCCGAAACCATG  
GACATGTGGCACGGCACCTGCACGTATCCGCGCGCCATTAAAGATGTGCTTGCCGACGTCGACCTCCTACACCGCGA  
AAACGTCAAGGTCGGCGCGCTTTCCGGAGGCGAACAACGACGCTTGATTGGCCTGCGCACTGCTTGCGGACCCCT  
CAATTTTGTTCCTCGACGAACCCACCACCGGCCTCGACCCAGAATCTAGGCGCCACACCTGGCAACTCCTGCTGGAC  
CTGAAACAGCGCGGCGTCAACCATGATGCTGACCACCCACTACCTGGAGGAAGCCGAATTCCTCTGCGACCGGATTGC  
CATCATGAACGCCGGTGAGATCGCAGTGGAAGGCACCTTGGATGAACTGGTGGCCCGCGAGAAGTCGATCATCAGTT  
TCGTGCTGCGTGGCGGGCAGGTGGAGTTGCCGGTCTTGAGTGGGGCTGAAATCATCCGCGACAACAACCACGTCCGC  
ATCGCCACCACCACCTGCAGCAGCACACCTTAGAAATACTTACCTGGGCTGCAGAGACCGGGATCGCGCTGGAAGG  
CTTCGCTGCAAAACCCGCCACCTTGGAATCCGTATTCATGGAC

>>RXA02095-amino acid sequence  
(1-1404, translated) 468 residues

MKTEQSQKAQ LAPKKAPEKP QRIRQLISVA WQRPWLTSFT VISALAAATLF ELTLPLLTGG AIDIALGNTG  
DTLTDDLDR FTPSGLSVLT SVIALIVLLA LLRYASQFGR RYTAGKLSMG VQHDVRLKTM RSLQNLDPG  
QDSIRTGQVV SRSISDINMV QSLVAMLPML IGNVVKLVLT LVIMLAISPP LTIIAAVLVP LLLWAVAYSR  
KALFASTWSA QOKAADLTTH VEETVTGIRV VKAFAQEDRE TDKLDLTARE LEAQRMR TAR LTAKFIPMVE  
QLPQLALVVN IVGGGYLAMT GHITVGTFVA FSSYLTSLSA VARSLSGMLM RVQLALSSVE RIFEVIDLQP  
ERTDPAHPLS LPDTPGLLSF NNVDFRGILN GFELGVQAGE TVVLVGPPGS GKTMAVQLAG NFYQPD SGHI  
AFDSNGHRTR FDDLTHSDIR RNLIADFDEP FLYSSSIPRE HLDGFGCQ

>RXA02095-nucleotide sequence A: upstream

CTCTCTTGGTCTCTCTCCCAACCCATTTTTAAGTACTCAAGACCCTTCCAACAGAAAGGATTACTCCCCCAACAGGCT  
CAAAAATACTGAAAGGCTCACGC

>RXA02095-nucleotide sequence B: coding region

ATGAAACTGAGCAATCCCAAAAAGCACAATTAGCCCCTAAGAAAGCACCTGAAAAGCCACAACGCATCCGCCAACT  
TATTTCCGTGGCGTGGCAGCGACCTTGGCTCACCTCATTACCGTAATCAGCGCTTTAGCTGCAACGTTGTTTGAAC  
TTACACTTCCTCTTTTGACCGGTGGCGCCATCGATATCGCGCTCGGAAATACCGGAGATACTTTAACCACTGACCTG  
CTGGACCGGTTCACTCCGAGTGGATTAAGCGTGTTGACCAGCGTCATTGCCCTTATCGTGCTTCTCGCGTTGCTTCG  
CTATGCCAGTCAATTTGGACGGCGATACACCGCAGGCAAGCTCAGCATGGGGGTACAGCATGATGTCCGGCTTAAAA  
CGATGCGCTCATTGCAGAACCTCGATGGGCCAGGTCAGGACTCTATTTCGCACAGGCCAAGTAGTCAGTCGGTCCATT  
TCGGATATCAACATGGTGCAAAGCCTTGTTGGCGATGTTGCCGATGTTGATCGGAAATGTGGTCAAGCTTGTGCTCAC  
TTTGGTGATCATGCTGGCTATTTCCCGCCGCTGACCATCATCGCTGCAGTGTGGTGCCCTTTGCTGTTGTGGGCCG  
TGGCCTATTTCGGAAAAGCGCTTTTTCGCTCCACGTGGTCGGCCAGCAAAAGGCTGCGGATCTGACCACTCATGTG  
GAAGAACTGTCACGGGTATCCGCGTGGTCAAGGCATTTGCGCAGGAAGACCGCGAGACCGACAAATTGGATCTCAC  
CGCACGTGAGTTATTTGCCAGCGCATGCGCACTGCACGTCTGACGGCAAAGTTCATCCCATGGTTGAGCAGCTTC  
CGCAGCTTGCTTTTGGTGGTCAACATTGTTGGCGGTGGCTATTTGGCCATGACTGGTCACATCAGGTGGGCACGTTT  
GTGGCGTTTTCTTCTATCTCACTAGCTTGTGCGCGGTGGCTAGGTCCCTGTGCGGCATGCTCATGCGCGTGCAGTT  
GGCGCTGTCTTCTGTGGAGCGCATCTTTGAAGTCATTGATCTTCAGCCTGAACGCACCGATCCTGCACACCCCTGT  
CACTTCCCGACACTCCCCTGGGTCTGTGTTCAACAACGTAGATTTCCGTGGGATTCTCAACGGTTTTTGAGCTGGGT  
GTTCAGGCCCGGTGAAACCGTTGTGTTGGTGGGCCCTCCAGGTTTCAAGCAAGACCATGGCTGTGCAGCTTGCTGGAAA  
CTTTTATCAACCAGACAGCGGCCACATCGCCTTTGATAGCAACGGCCATCGCACTCGCTTCGACGACCTCACCCACA  
GCGATATCCGCAGGAATCTCATCGCGTTTTTGTATGAGCCGTTCTTGTACTCCTCCTCCATACCGCGAGAACATCTC  
GATGGGTTTGGATGTCAG

>RXA02095-nucleotide sequence C: downstream

TGATGAGCAGATCGAACACGCAG

>>RXA02074-amino acid sequence

(1-1623, translated) 541 residues

MRSLLRDIPA	VGWLITATIV	VRTLVLVALVI	VGIGLLIDVP	SPAHSAMLWW	VLGATAAAA	LLCAEAVLPQ
RIRARVERSW	RRQLAAKNLE	LNSSSSDDAQ	LITLATEATS	KASTYTMVFL	GPYFAVFLAP	LTVIAVVGAA
ISWPIAGILC	LGLCVIPFVI	SWAQRMLKGA	GAGYGRASQ	LAGVFLESVR	TLGTTMMLNA	AGORROIITQ
RAENMRSQVM	SLLYRNQIMI	LVTDGVFGVA	TTMVAAVFAI	GGFFSGSLTL	GQAVALLVLLA	RLRIDPINRM
GRTFYTMAG	KPSLIAIEKA	LATTFTDQPT	QQGQRHGDGL	VVNNLKIARD	HRDIVHGISE	SIPRGSHIAV
VGPSGAGKSS	VALALSGLLE	FDGAISLGGH	NCEMLDLRAS	VSFVPQSPTL	FSGSIKSNID	LARTGVDS DH
IHAALLGEEL	PADLKVGETG	KGVS GGQAAR	ISIARGLVKN	AAVIVLDEAT	AQLDYTNARQ	VRHLAKSLEC
TLVEITHRPS	EALDADFIIV	LEDGQLTMD	TPSNVSQHNA	FFRTAVMEEE	Q	

>RXA02074-nucleotide sequence A: upstream

CGGGGGAAGGCCGTGTCGCATGCTCGGGCTAGCCTTGGATCTCAAGAAGAATTCGACTGGTTTAAAGTCTGGGCTTT  
AAGTGCAGAAAGGTTGTGGATTG

>RXA02074-nucleotide sequence B: coding region

ATGCGCTCCCTGCTTCGTGATATCCCTGCGGTGGGTTGGCTAATCACCGCGACGATTGTTGTGCGCACGCTCGTTGT  
TGCGCTGGTCATCGTTGGGATCGGCTTGCTTATCGACGTCCCCTCGCCGCTCATTAGCCATGTTGTGGTGGGTTTC  
TGGCAGGTGCCACGGCAGCAGCTGCGCTGCTGTGCGCGGAAGCGGTGCTCCCCAACGTATTCGTGCACGAGTTGAA  
CGATCCTGGCGGGCGGCAGTTGGCTGCTAAAAATCTGGAGCTGAATCCAGTTCGTGATGATGCCCAGTTGATCAC  
ACTGGCAACTGAAGCCACCTCAAAGCATCCACTTACACAGTGATGTTTCTGGGGCCTTACTTTGCAGTATTTTTGG  
CCCCACTGACAGTTATTGCGGTTGTGCGCGCGGCTATTTCTGGCCGATTGCGGGGATACTGTGCCTCGGGTTGTGC  
GTGATACCTTTTCGTTATTTCTTGGGCACAGCGCATGTTGAAAGGCGCTGGCGCGGGATACGGGCGAGCATCTGGGCA  
GTTGGCAGGCGTGTTTTTGAATCGGTGCGCACACTAGGCACCACGATGATGCTGAATGCCGCTGGGCAGCGCAGGC  
AGATCATCACACAGCGCGCAGAGAATATGCGCTCCCAAGTGATGTCATTGCTGTACCGAAATCAGTTGATGATTCTG  
GTGACCGACGGCGTGTTTGGAGTTGCCACCACAATGGTTGCTGCGGTGTTTGCCATTGGAGGATTCTTTTTCAGGCTC  
TCTTACTCTCGGCCAAGCTGTAGCACTCGTATTGCTGGCCAGGCTGCTTATTGATCCCATCAACCGCATGGGTGCGA  
CGTTTTTACACCGGCATGGCAGGCAAACCTCGCTGATCGCCATTGAAAAAGCCCTCGCGACAACCTTTACTGATCAG  
CCAACTCAACAGGGACAGCGCCACGATGGGGATCTGGTGGTCAACAACCTTGAAGATCGCCGCGATCACAGGGACAT  
TGTGCACGGTATCTCTTTCAGCATTCCCCGCGGTTCCACATCGCGGTGGTAGGTCCAGTGGCGCTGGTAAATCCT  
CTGTGGCTCTAGCGTTGTCCGGACTTTTAGAGTTTGATGGTGCGATTTCCCTCGGCGGCCACAACCTGTGAGATGTTA  
GATCTTCGCGCCTCAGTCAGTTTCGTGCCCAATCCCCACGCTGTTTAGCGGAAGCATCAAAGCAATATCGATCT  
GGCGCGCACGGGTGTTGATTCTGATCATATCCACGCAGCACTTTTAGGCGAAGAACTCCCCGCGGACCTCAAAGTCG  
GTGAAACCGGCAAAGGTGTCTCCGGCGGCCAAGCAGCACGCATTTCCATTGCCCGAGGTTTAGTAAAGAATGCTGCC  
TGATTGTTCTCGACGAGGCGACCGCACAACTCGACTACACCAACGCCCGCAGGTTGCACATCTTGCCAAATCCCT  
TGAGTGACAGTTGGTTGAGATCACCCACGCGCCATCAGAAGCCCTCGATGCAGACTTCATCATTTGTTTTAGAGGATG  
GCCAATTGACCATGATGGATACACCCAGCAACGTTTTCCAGCACAAATGCGTTTTTTCGCACCGCTGTGATGGAGGAA  
GAACAA

>RXA02074-nucleotide sequence C: downstream

TGATTTCCTCGACTTCTCCAATTG

&gt;&gt;RXA02224-amino acid sequence

(1-1797, translated) 599 residues

MAQHERVADA LQPASLAESW RELKTMPSGP KAWWYVSFV ISVVTVMAMV GTSNLLGYSV DLINGQSLPL  
IGSGSTAMIW LLGLVGAGIL AETAGRALLQ LVINTLARRL SVDLRKAALS SALRAPVPDV MELGTGNVIS  
RLTQDIDNTV RIVGMVGURL VITILILPSS LFALMTIHWT FVILFIAVIV VLIPSGRKAV RAIPSATNIV  
SSTEARRNNL LLDTIRGIET LRVLKLGAWG VQMRQASWT AVQATADRAP IFTRLLALGS IAYGLLLIGV  
FGLSAFWVAQ DAMSIGAATA AVFVVVRMEI HVFNVLFFAS EIQSASTSLG RAVSLAQMAR RTEQLSESAD  
CTEPPSVTVQ DVTFKYPGGV AILEDENLVL EAGTTTALVG TSGAGKSTLA GVIAGLQRPD SGAVLVGGIN  
TATVTDTWTT RQVALISQEV HLFAGTLAED LRMANAHATD AQLHAALESV GLGQMTTAFR RFFPSGLDTK  
IGAGAEELTP EIQQQISLAR IVLRNPPVLI MDEATSEAGS DDARMLEKAA TEIARNRTTL VVAHRLDQAV  
VADRIIVMEQ GTITEDGTHQ ELLAFEGRYA QLYQRWSAQ

&gt;RXA02224-nucleotide sequence A: upstream

GCTTCGTCGAGGCCGAAAACCGTCGTCATTACGTCGAACCCGACGTGGCACGGCGTCGCAAAGCAGATGCAATCTG  
ATTTTTCGGAAGGGGTGAAGTAG

&gt;RXA02224-nucleotide sequence B: coding region

ATGGCGCAGCATGAGCGCTTGCGGATGCGCTGCAGCCGGCGTCGTTGGCGGAGTCGTGGCGTGAGCTGAAAACGAT  
GCCTTCGGGGCCCAAGGCCTGGTGGTATGTGAGTTTCGTGGTTATTAGCGTGCTACGGTCGTGGCGATGGTCGGCA  
CGTCCAATTGTTGGGCTATTCGGTTGATCTGATCAATGGGCAGTCGTTGCCGCTGATCGGTTTCAGGATCGACCGCA  
ATGATCTGGTTGGTTGGTTGGTGGCGCTGGAATTTAGCAGAAACGTCGCGGTGCGCGCTGCTGCAATTGGTGAT  
CAACACCTTGGCACGTCGCTGTGCGTGGATCTGCGGAAAGCTGCGCTGTCTTCGGCGTTGCGTGCAACCGGTTTCCTG  
ATGTGATGGAATTGGGCACGGGAAACGTGATTAGCCGCTGACGCAAGACATCGATAACACTGTGCGCATCGTCGGC  
ATGGTAGGTGTGCGTTTGGTGATCACCATTGTTGATTCTGCCCAGCTCCTTGTTCGCGTTGATGACCATTCACTGGAC  
CTTTGTGATCCTGTTTCATCGCAGTGATTGTGGTGCTGATTCCAGCGGTGCGAAAGCCGTGCGAGCTATTCTTCCTCGG  
CAACAAATATTGTGTCCAGTACGGAGGCGCGTCGAAACAACTGCTCCTCGATACGATCCGTGGCATTGAAACACTG  
CGTGTGCTCAAGCTCGGTGCGTGGGGTGTGACGCGGATGCGCCAAGCGTCGTGGACTGCGGTGCAAGCAACAGCTGA  
TCGCGCGCCGATTTTCACTCGTCTGCTCGCCCTTGGTTGCTTATGGCCTGCTGCTAATTGGCGTGTGTTGGGC  
TCAGTGCGTTTTGGGTTGCCCAGGATGCGATGAGCATTGGAGCGGCAACGGCAGCAGTTTTCTGTGGTTGTGCGCATG  
GAAATTCACGTGTTCAACGTGCTGTTCTTCGCATCGGAAATTCAGAGTGCCTCTACTTCTCTTGGTCGCGCGGTGTC  
CCTTGCCCAGATGGCTCGTCGCACCGAACAGCTGTCTGAGTCTGCCGATTGCACAGAACCACCTCCGTGACTGTGC  
AGGACGTGACGTTTAAATATCCCGGCGGCGTGGCCATTTTGGAGGATTTCAATCTGGTCTTGAAGCAGGAACAACC  
ACAGCGCTGGTGGTACTTCTGGTGCGGGAAAATCCACGCTCGCGGGCGTCATTGCGGGGCTGCAGCGCCCTGATTC  
CGGCGCCGTTTTGGTGGGGGTCATCAACACCGCCACCGTCACCGACACGTGGACTACCGCCAGGTTGCGCTGATCA  
GCCAGGAAGTCCACCTTTTCGAGGCACTCTGGCAGAGGATCTTCGCATGGCCAATGCGCACGCCACCGACGCGCAG  
CTCCACGCAGCGCTCGAGTCGGTGGGCTCGGGCAAATGACAAGTCTTTTCGACGTTTCTTTCCATCCGGATTAGA  
CACCAAAATTGGCGCCGGCGCAGAAGAACTCACCCCTGAAATCCAACAGCAAATCTCTCTTGGCCGCATCGTGCTCC  
GCAATCCACCTGTGTTGATCATGGATGAAGCCACCACTGAAGCCGGCAGCGATGATGCCCGCATGTTGGAAAAAGCC  
GCCACAGAAATCGCACGAAACCGCACCACTTGGTTGTGCGCACCGCCTTGACCAAGCAGTTGTGCGCAGATCGCAT  
CATCGTGATGGAACAAGGCACAATCACCGAAGACGGCACTCACCAGGAATTACTTGCTTTTGAAGGGCCGCTACGCGC  
AGCTGTATCAACGATGGAGTGCTCAA

&gt;RXA02224-nucleotide sequence C: downstream

TAGTTCAAATCCACCACAACTC



>>RXA01947-amino acid sequence

(1-279, translated) 93 residues

ILLDEPTNDL DVETLGSLN ALQNFPGCAV VISHDRWFLD RTCTHILAW E GNIAEGQYW FEGNFEDYEK  
NKVERLGADA ARPSRVTHRK LTR

>RXA01947-nucleotide sequence B: coding region

ATCCTCCTCGATGAGCCTACAAACGACCTCGACGTGGAAACTCTGGGCTCCCTGGAAAATGCACTCCAGAACTTCCC  
TGGCTGTGCAGTGGTCATTTCCCACGACCGTTGGTTCCTGGACCGCACCTGTACCCACATCCTCGCGTGGGAAGGCA  
ACATCGCGGAAGGCCAGTGGTACTGGTTCGAAGGCAACTTCGAAGACTACGAGAAGAACAAGGTGGAGCGCCTCGGT  
GCTGACGCCGCACGTCCTTCCCGTGTCACCTCACCGCAAGCTGACTCGC

>RXA01947-nucleotide sequence C: downstream

TAGAGTCGAAGCAGCGCTAAACC

>>RXA01946-amino acid sequence  
(1-1275, translated) 425 residues

IRKYSRLEEQ FQSLGGYEAD AEAAQICDNL GLEARILDQQ LKTLSGGQRR RVELAQILFA ATNGSGKSKT  
TLLLDPEPTNH LDADSITWLR DFLAKHEGGL IMISHDVELL GAVCNKIWYL DAVRSEADVY NMGFSKYVDA  
RALDEARRRR ERANAEEKKAG ALKDQAARLG AKATKAAAK QMIARAERMI DNLDEIRVAD RAANIVFPEP  
APCGKTPLNA KGLTKMYGSL EVFAGVDLAI DKGSRVVVLG FNGAGKTLL KLLAGVERTD GEGGIIVTGYG  
LKIGYFAQEH DTIDPKSVW QNTIEACADA DQSLRSLLG SFMFSGEQLD QPAGTSLSGGE KTRLALATLV  
SSRANVLLLD EPTNNLDPIS REQVLDALRT YTGAIVLVTH DPGAVKALEP ERVIVLPDGT EDLWNDQYME  
IVELA

>RXA01946-nucleotide sequence B: coding region

ATCCGCAAGTACTCCAGGCTCGAGGAACAATTCCAGTCGCTCGGCGGCTACGAAGCTGACGCCGAAGCAGCCCAGAT  
CTGCGACAACCTCGGCCTCGAGGCACGCATCCTCGACCAGCAGCTTAAACCCCTGTCCGCGCGCCAGCGCCGCGCGG  
TCGAGTTGGCGCAGATCCTCTTCGCGCCACCAACGGCTCCGGCAAATCAAAAACCACATTGCTTCTCGACGAGCCC  
ACCAACCACCTTGGACGCAGACTCGATCACCTGGCTCCGTGACTTCCTGGCGAAGCACGAAGGTGGACTGATCATGAT  
TTCGCACGACGTCGAACCTGCTTGGCGCCGTATGTAACAAGATTGGTACCTCGACGCAGTACGCAGCGAAGCCGATG  
TCTACAACATGGGCTTTAGCAAATACGTCGATGCACGTGCACTCGATGAAGCACGCCGACGCCGTGAGCGCGCAAAC  
GCCGAAAAGAAGGCCGAGCCCTCAAGGACCAGGCTGCACGCCCTCGGCGCGAAAGCAACCAAGGCTGCCGCAGCTAA  
GCAGATGATCGCCCGTGCAGAACGAATGATCGACAACCTCGACGAAATCCGCGTAGCTGACCGCGCCGCCAACATCG  
TTTTCCCGAGAACCAGCACCCCTGTGAAAAACCCCACTCAACGCCAAGGGCCTGACCAAGATGTACGGCTCCCTCGAA  
GTCTTCGCGCGCGTTCGACCTAGCCATCGACAAAGGCTCCCGCGTAGTCGTCTCCTCGGATTCAACGGTGCAGGTAAAC  
CACCCCTGCTCAAACCTCCTCGCCGGTGTGGAACGCACCGACGGCGAAGGCGGCATCGTCACCGGATACGGCCTCAAAA  
TCGGCTACTTCGCCCAGGAACACGACACCATCGACCCCGACAAATCCGTCTGGCAAAACACCATCGAAGCCTGCGCC  
GACGCCGACCAACAAAGCCTCCGCAGCCTCCTCGGATCCTTCATGTTCTCCGGCGAACAACCTCGACCAACCAGCAGG  
AACACTCTCCGGCGGTGAAAAAACCCGCTCGCACTGGCCACCCTCGTGTCTCTCCCGCGCAACGTCCTGCTTCTCG  
ACGAGCCCACCAACAACCTTGACCCGATCTCCCGCGAACAGGTCTCGACGCACTGCGCACCTACACCGGCGCAGTC  
GTCCTGGTTACCCACGACCCGGGTGCAGTCAAGGCCCTTGAGCCAGAACCGGTCATCGTGCTTCCTGATGGCACC  
GGATCTTTGGAATGATCAGTACATGGAAATCGTGGAATTGGCG

>RXA01946-nucleotide sequence C: downstream  
TAGGTTCTAAGGCTGTTTATGCT

>>RXA01881-amino acid sequence  
(1-441, translated) 147 residues

MANLINLENV SKTWGLKTL DGVSLGVQTG DRIGVVGLNG GGKTTLLEVL TGIEKPDQGR VSHNSDLRMA  
VVTQRAELND DDTVADVVLG PLGLEVFWEA SNATVRDVLG GLGIVDLGLD TKVGQTFSGG RSADAPTWPP  
RWFATLT

>RXA01881-nucleotide sequence A: upstream

ACCGGCCCTGCGGCCTCAACCGCCGACCAGCGCGCGCACACATTTGACTGTTTCATAATAAAGACAACTTAAGT  
ATCGGAGTCGAAGAAAAACCACA

>RXA01881-nucleotide sequence B: coding region

ATGGCCAATCTGATTAATCTCGAGAACGTCTCCAAAACCTGGGGATTAAAAACGCTTCTCGACGGTGTCTCCTTAGG  
TGTTCAAACCGGCGACCGCATTGGCGTCGTCGGCCTCAATGGTGGCGGAAAAACCACCCTGCTGGAAGTACTTACTG  
GCATCGAAAAGCCGGATCAGGGCCGTGTGTCTCACAACCTCTGACCTGCGCATGGCTGTGGTGACGCAGCGTGCTGAA  
CTCAATGATGACGACACCGTCGCTGACGTGGTGCTTGGACCTTTGGGTTTGGAAGTTTTCGAATGGGCATCAAACGC  
CACGGTGC GCGACGTCTCTGGTGGCTTGGGCATTGTCGATCTTGGCCTTGACACCAAGGTGGGGCAAACCTTTCCG  
GTGGGCGAAGCGCCGACGCACCAACCTGGCCGCCGCGCTGGTTCGCGACCTTGACC

>RXA01881-nucleotide sequence C: downstream

TGATCGTGCTCGACGAGCCCACC

>>RXA02253-amino acid sequence

(1-927, translated) 309 residues

MIQSTGVTHT DKSAQENPVK YRDNFTPVII TGMSGAGLST AARVLEDLGW YVAHNIPPQI ILELIDMCAR  
EDSPVDKVAV VCDVRSREFR GSILTQVSEL RDKQLDPTVL FLEARDEVLI KRFDNVRRTTH PLQGSQTLQV  
GIERERTVLS PVKEDASVVI DTSDLVHDL RRAIESSFRT IATRTQHVTI ESFGFKHGSP RDADEFVVDVR  
FLPNPFWVPE LRPFRGVDPK VSDYVLSQKG AEEFLNNFVD MLKDMLEPGYR HEGKNFITIG VGCTGGHHRS  
VAVSEELAKR IADQTTLDVS VVHRDINRH

>RXA02253-nucleotide sequence A: upstream

TGAGGACATTTCCCAAGTGAAAGGCTTCGGCCCGAAACTTGCGGAGGCTGTCTATGAAGGTCTTCACGCGTCAAAAT  
AAGTAGATCGCTAGGATGTAACC

>RXA02253-nucleotide sequence B: coding region

ATGATTCAATCCACTGGGGTCACGCACACTGATAAGTCTGCACAAGAAAATCCTGTGAAGTACAGGGACAATTTAC  
ACCTGTCATCATCACCGGTATGTCAGGCGCAGGTCTGAGCACAGCAGCCCCGAGTCCTCGAAGACTTGGGTTGGTATG  
TGGCGCACAATATTCACCACAGATCATCCTGGAATTAATTGATATGTGCGCTCGGGAAGATTCTCCCGTCGACAAA  
GTTGCAGTCGTGTGCGATGTGCGCTCCCGTGAATTCGCGGAAGCCTCACCCAGGTTGTTTCAGAGCTGCGTGATAA  
GCAGCTCGATCCCACGGTGTTATTTCTGGAAGCACGCGATGAGGTGCTGATCAAGCGATTGATAATGTGCGCCGCA  
CCCATCCTTTGCAGGGCAGCCAAACCCCTTCAGGTGGGTATTGAACGTGAACGAACCGTGCTGTCTCCTGTGAAGGAA  
GACGCTTCAGTGGTCATCGATACCTCGGATCTGTCCGTGCATGATTTGCGCCGCGCCATCGAATCCTCGTTTAGGAC  
AATCGCCACGCGCACCCAGCACGTCACCATTTGAATCATTCCGTTTCAAACACGGCTCACCCACGCGACGCCGACTTTG  
TTGTGGACGTGCGTTTCTTCCGAACCCATTCTGGGTTCCAGAGCTGCGCCCATTAGGGGAGTGGACAAGCCAGTA  
TCTGACTATGTGCTCTCCCAAAAAGGCGCAGAAGAATTTTGAACAACCTTTGTGGACATGCTCAAAGACATGCTTCC  
GGGATACCGCCACGAAGGAAAAAATTCATCACAATCGGTGTCGGCTGCACCGGTGGACACCACAGATCAGTTGCGG  
TGTCTGAAGAACTAGCCAAAAGAATCGCAGATCAGACCAGCTCGACGTGTCTGTAGTACACCGCGATATTAACCGC  
CAC

>RXA02253-nucleotide sequence C: downstream

TAGGAAAGGGGCCAACTAATTGA

>>RXA02225-amino acid sequence  
(1-882, translated) 294 residues

QTEERFGAAA DEALAIMLKE ARLQSLTFV RQLVPAVFSV GLLAYASLLA FDGDITGGEM ISVTLLVPPS  
LTVLGVSLGM MTEIWARGQA STKRQNLVT ELDKAAAEPR PQPATFEFEE GITVWDPSTP EARDVIDREL  
EALQVREDVI VAPHRVSVFE GVLKDNLNPM GTIAPEMLRA ALHAASCEDE LSRLGADLNM PGEFELPDTL  
IGEAGLNLSG GQRQRIALAR FLAVDPVELI LDEPTTGLDA VTILDEVHRV EKLRRGRKTV VITSNPTWHG  
VAKQMQSDFS EGVK

>RXA02225-nucleotide sequence B: coding region

CAAACGTAGGAGCGCTTTGGCGCAGCGGCTGATGAAGCCTTGGCAATCATGTTGAAGGAAGCTCGTCTGCAGTCGTT  
GCTGACTTTTGTGCGCCAACCTGTCCCAGCGGTGTTTTCTGTGGGTCTTTTGGCTTATGCGTCACTGTTGGCTTTTG  
ACGGTGACATAACTGGTGGTGAGATGATCTCGGTGACGTTGCTGGTGCCACCTTCGTTGACTGTGTTGGGTGTGTGCG  
CTTGGCATGATGACAGAGATTGGGGCTAGGGGACAGGCTTCGACAAAAAGGGTCCAAACTTAGTCACTGAAGTGA  
TAAGGCGGCGCTGAGCCACGACCTCAGCCTGCCACCTTTGAATTTGAAGAGGGGATCACGGTGTGGGATCCTTCGA  
CACCTGAGGCACGCGATGTGATTGATCGGGAGTTGGAGGCGCTTCAGGTTCCGGAAGATGTCATTGTGGCTCCTCAC  
CGCGTCAGCGTGTTTGAAGGTGTGCTGAAGGATAATTTGAATCCGATGGGCACTATCGCACCGGAGATGCTGCGCGC  
TGCTCTTCATGCTGCAAGTTGTGAGGACATCTTGAGCCGATTGGGTGCTGATCTGAACATGCCGGGGGAGTTTGAGC  
TTCCAGATACCTTGATCGGCGAGGCCGGATTGAATCTCTCCGGTGGCCAACGCCAGAGGATTGCTTTGGCACGATTC  
TTGGCTGTTGATCCTGAGGTGCTCATTTTGGATGAACCGACCACGGGGTTGGATGCGGTGACCCCTGGATGAAGTGGC  
ACATCGCGTCGAAAAGCTTCGTGAGGCCGGAACCGTCGTACGTGCAACCCGACGTGGCACGGCGTCGCAA  
AGCAGATGCAATCTGATTTTTTCGGAAGGGGTGAAG

>RXA02225-nucleotide sequence C: downstream

TAGATGGCGCAGCATGAGCGCGT

>>RXA02749-amino acid sequence  
(1-876, translated) 292 residues

MSPILKVRDL VKRYGDTVAV DGLNFDVSQG EIFAFLGENG AGKTTTISCL IGIDQATSGE IELQGGQVDS  
EKLGVVFQQS VLDPLLSAKE NLETRGQLYP GVGKQ RVAQL IEQIGMEGFA DRRYGVLSGG EKRRTDIARA  
LLHSPDILFL DEPTAGLDPR SRRQVWD TIN SLRNDVGLTV FLTHYMEET ELADSVLIID RGKEVASGTP  
MELRARYTTT ELTLRTNDPT HSGKELAHLS PEIDGDLRI KLENGLEAAR LATELDGVLD VEIRHGSMD  
VFLAVTAERK RS

>RXA02749-nucleotide sequence A: upstream

CAACCTAGACTTCGGTAAGAAGTAACTTTGCTTTAGTTGGTCGGCGCATCACTTTCCCTAAGCGATGCGCCGATTAC  
TTGTTTTTGTCTACAAATTTAACT

>RXA02749-nucleotide sequence B: coding region

ATGTCACCCATCCTAAAAGTGC GGGACCTCGTCAAACGCTATGGCGACACCGTTGCGGTTGACGGTTTAAATTTTGA  
TGTTTTACAAGGGGAAATTTTTGCCTTTCTAGGGGAGAACGGCGCAGGAAAAACAACCACGATTTTCATGCCTGATTG  
GCATTGATCAAGCCACCTCTGGGGAGATCGAACTGCAGGGTGGCCAAGTAGATTCTGAAAACTTGGAGTGGTGT  
CAACAATCCGTCTTAGACCTTTGCTGAGTGCCAAAGAAAAGTTGGAAACACGCGGACAGCTGTACCCAGGGGTGGG  
GAAGCAGCGGGTTGCACAGCTCATTGAGCAAATCGGGATGGAAGGGTTGCGGACCGCCGATACGGAGTGTGTGCGG  
GCGGTGAAAAACGTCCGACCGACATCGCAGGAGCTTTACTGCACAGCCAGACATTCTTTTTCTTGATGAACCCACA  
GCAGGCCTCGACCCAGATCACGACGCCAAGTTTGGGACACCATCAATCCCTGCGTAACGATGTGGGCCTCACTGT  
CTTTTTGACCACTCACTACATGGAAGAAACAGAACTGGCTGATTCAATTCTAATCATTGACCGTGCGAAAGAGGTG  
CATCAGGAACCCCGATGGAAGTGC GCGCCCGTTACACCACAACAGAATTGACTCTTAGAACAAACGACCCTACTCAT  
TCGGGTAAAGAGTTGGCCCACTTGAGCCAGAAATCGACGGTGACCGACTGCGGATCAAGTTGGAAAATGGGCTCGA  
AGCTGCGCGCCTGGCAACAGAACTAGATGGGGTTCTCGACGTAGAGATCCGCCACGGTTCATGGACGATGTATTT  
TAGCAGTTACAGCTGAACGGAAACGATCA

>RXA02749-nucleotide sequence C: downstream

TGATTACAGTTCTGACACGCAGA

>>RXA02571-amino acid sequence

(1-1029, translated) 343 residues

VVALTQIVGP SGSGLTRELE KRYRETPGAV MLTADPRAHI TYLRATVAEE LAFGLEQIRGI VPAQMWERVR  
NIGLGLENLL DRAPAQLSGG QTRRLAIGTV AILEAPTMLL DDPLSGLDTS SRAQLITMLE SYEGDVIVAA  
HKRWLDAPTV YLGDLEELSL PARVEFSGPS RTFSAITGTR GQRRRWQF NESQPQFQIG PLDITVSAGO  
VLWLQGPNGS GKSTLLRGLA NEPGTEIMLQ NPSDQVIDST VANWVPGSNS EEHPLDLSQR ELRLAQCDAA  
LGNNEVLLA DEPDVGLDVG GRNAIHQFRA DFLNGGALI LTCHDETIVA EVAEYAIKVE MGL

>RXA02571-nucleotide sequence A: upstream

TGGACAGGCCGCGGCGCGGTACGGTGTGGTTGAGGTGGTGGAGGGCGCGCTCGAAAAGCATTGTCGCTGGTTGTTG  
CCGCTTTTGGCAGTCGGGATGGC

>RXA02571-nucleotide sequence B: coding region

GTGGTGGCTCTAACTCAAATCGTCGGACCGTCCGGCTCCGGGCTCACGCGGAATTGGAAAAACGCTACCGGGAAAC  
GCCCCGAGCGGTGATGCTGACCGCCGACCGCGCGCATATCACCTACCTGCGCGGACAGTCGCCGAGGAGCTGG  
CCTTTGGGCTGGAACAACGCGGCATCGTACCCGCGCAGATGTGGGAGCGCGTCCGAAACATCGGGCTCGGCCTCGAG  
AATCTGCTAGACCGCGCACCCGCGCAACTTTCCGGCGGGCAAACACGGCGGCTGGCGATCGGCACCGTCGCCATCTT  
AGAGGCGCCAACGATGCTTCTCGACGACCCCTCTCCGGTCTTGATACCTCCTCGCGAGCCCAACTCATCACAATGT  
TGAATCATATGAGGGCGATGTCATCGTCGTCGCGACAAGCGGTGGCTCGACGCGCCGACTGTGTACTTAGGGGAT  
TTGGAGGAGCTGTCCCTGCCTGCGCGGGTGGAAATTTCCGGTCCATCGCGAACGTTTTTCAGCGATTACAGGAACCCG  
CGGACAACAACGCGGACGCTGGTGGCAATTCAACGAATCCCAACCACAGTTTCAGATCGGCCCCCTGGATATTACTG  
TTTCTGCAGGTCAAGTGCTGTGGTTGCAGGGTCCCAATGGTTCAGGGAAGTCCACACTCCTGCGTGGTCTTGCCAAT  
GAACCCGGCACTGAATTGATGCTGCAAAACCTAGCGATCAAGTCATTGACTCCACTGTTGCTAATTGGGTGCCAGG  
CAGTAACAGTGAAGAACATCCGCTGGATTTATCGCAACGCGAACTCCGCCTTGCCCAATGCGACGCAGCCCTGGGTA  
ATAACCCGGAAGTTTTTGCTTGCTGATGAACCCGACGTCGGCCTTGATGTGCGCGGTGCGAAACGCCATCCACCAGCGC  
TTTGCGGATTTCTTAGGGAATGGGGGAGCGCTGATCCTGACCTGCCATGATGAAACCTTCGTGCCAGAGGTAGCTGA  
ATACGCGATAGTGAAGGAAATGGGGCTC

>RXA02571-nucleotide sequence C: downstream

TAGGTTTCTTTGGACCAACCAC

>>RXA02547-amino acid sequence  
(1-2124, translated) 708 residues

AARLTVDDEYP AAREALESAG QRNVEDRTRA VDEFKAADQE LSSLSKGSSN IEYRLLQVRE NLCQDLGVSP  
RDMPPFAGELI DPNNAEWEFV VQRILGGFAA EMLVPHGLLP RVRDWNNAKH LAALLKFNGV VTTGEYKTSR  
FPADSLIRKV DVVESPFERDWN VNQELGKRFN IRCVRTPEEL SALGPRDQGV TILGVRKFAQ QTGDPTTRWE  
KDDRRKLGDR STYRLGSTND AKVETLRETV KAGKAVVQAA DNRIANRAE LRELERQYQA SQEILKVSWA  
QIDVESADAA IAEIDRLLEE LNNTPEATEL SARHEAAKQT LARVSDLLVA AQSEETVASM NLKRAETELK  
RLESPLVAEV SEEIAREVEK LFLANTRRVH AANVDEQTA LREDLDKQID ANEAE LRRCENQIVGILRSY  
IETWPANRAD LQAEPEFVGE AINRLGELRS DRLAEFTAKF LGLMNMSTR NLGQISRRLR DARREIEERI  
EPINASLAQS EFNENRFLHI DIRDQSGPIV REFQOKLDA TSGDILGTSTE KQAFARYALI AEIISKLASH  
DSADARWRNT VLDTRRHVRF IGLERDSGA TVNTYVDSAS LSGGQAQKLV FFCLAAALRY QLAEPGAHY  
TYATVILDEA FDRADPAFTR QTMNVFHSFG FHMVLATPLK LIQTLGDYVG STIVVSYTEK PNAQGAIQGN  
SSFSRIEK

>RXA02547-nucleotide sequence B: coding region

GCTGCGCGGCTGACCGTGGATGAGTATCCGGCGGCGAGGGAAGCGCTTGAATCTGCAGGTCAGAGGAATGTAGAGGA  
CCGAACCCGTCGGTTGATGAGTTCAAAGCGGCGGATCAAGAGCTGTCTTCTTTGAGTAAAGGCAGCAGTAATATTG  
AGTACCGTTTTGCTGCGAGGTGCGGGAATTTGTGTGTCAGGATTTGGGCGTGAGCCCGCGGATATGCCCTTTGCCGGT  
GAGCTGATTGATCCGAATAATGCGGAATGGGAACCGTTGTGTCAGCGCATTTTGGGTGGTTTTGCTGCGGAAATGTT  
GGTTCCTCATGGGTGTTGTGCCACGGGTTCCGGGATTTGGGTAAATGCCAACATTTGGCAGCGCTGCTGAAATTCAACG  
GCGTGGTGACAACGGGGGAGTACAAAACCTCGCGTTTTCCGGCGGATTCCTGATCCGAAAAGTTGATGTTGTGGAG  
TCGCCGTTTTCGCGATTGGGTAAATCAAGAATTAGGCAAGCGTTTTTAATTCGGTGGTGCGCAGTCCCTGAGGAATT  
GTCGGCGCTGGGGCCACGCGATCAGGGCGTGACCATTTTGGGTGTGCGAAAATTTGCGCAGCAGCAGCGCATCCGA  
CGACGCGTTGGGAAAAAGATGATCGCCGAAAGCTGGGGGATCGTTCCACATACCGTTTTGGGTTCACCAATGATGCC  
AAGGTGGAACGCTTCGGGAAACCGTGAAAGCTGGCAAAGCAGTTGTGTCAGGCAGCTGATAATCGCATTGCTGCAAA  
CCGCGCTGAGCTGCGGAACTTGAACGGCAGTATCAAGCTTCGCAAGAAATTTTGAAAGTGTGCTGGGCTCAGATTG  
ATGTGGAATCAGCCGACGCGGCGATTGCTGAGCTGGACCGATTGCTGGAAGAGCTGAACAACACTCCAGAGGCCACC  
GAGCTTTCCGCGCGGCATGAGGCGGCGAAGCAGACGCTCGCGAGGGTTTTCTGACTTGCTTGTGTCAGCTCAGAGTGA  
GGAAACCGTGGCGTCGATGAACCTGAAACGCGCGGAACTGAATTGAAACGGCTCGAAAGCCTGCCGGTTGCGGAGG  
TTTCTGAAGAAATCGCGCGGGAAGTGGAGAACTATTTCTTGCCAACACCCGCGGGTTCACGCCGCCAACGTTGAT  
GAGCAGACCATTTGCGCTGCGCGAGGATCTGGACAAACAAATCGATGCCAATGAGGCAGAACTTCGACGTTGTGAAAA  
CCAAATTGTTGGCATTGTTGCGCAGCTATATTGAAACGTGGCCTGCGAACCGCGCTGACTTACAAGCCGAACCTGAGT  
TTGTTGGTGAGGCCATCAACCGCCTCGGCGAGCTTCGCGAGCGATCGTTTGGCAGAAATTCAGGGCCAAATTCCTAGGG  
CTCATGAACGAGATGTCCACCCGAAACCTCGGCCAAATCTCGCGGCGTCTACGTGATGCGCGCCGGGAAATCGAGGA  
GCGCATCGAGCCGATCAACGCCTCCTTGGCGCAGTCGGAATTCACGAAGGTCGCTTCCTGCACATCGACATCCGTG  
ATCAAAGTGGTCCGATTGTGAGGGAATTCAGCAGAACTTGATGCCGCTACCAGCGGTGACCTGGGAACAGTACC  
GAGAAACAAGCCTTCGCCCGTTATGCGCTGATCGCTGAAATCATTTCCAACTCGCCTCCCAGACTCCGCCGACGC  
CCGCTGGCGCAACACCGTTCTAGACACCCGCCGCCACGTTTCGCTTCATCGGCCTCGAGCGGATTCGACGGCGCAA  
CCGTCAACACCTACGTGCACTCCGCATCACTTTAGGCGGACAAGCCCAGAAGCTGGTGTGTTTTCTGCCTCGCCGCT  
GCCTTGCGCTACCAGCTAGCCGAACCCGGCGGCCATTATCCACCTACGCCACCGTCATTCTGGACGAAGCCTTCGA  
CCGCGCCGACCCCGCCTTCACCCGCCAAACCATGAACGTCTTCACAGCTTCGGCTTCACATGGTGTCTCGCGACCC  
CGCTGAACTTATCCAAACCTCGGCGATTATGTGCGCTCCACCATCGTGGTCAGCTACACCGAAAAACCAACGCC  
CAGGGCGCAATTCAGGGCAATTCAGTTTCTCTAGGATCGAGAAA

>RXA02547-nucleotide sequence C: downstream  
TAACATGCCATTGTTTATCGACG



>>RXA01604-amino acid sequence  
(1-606, translated) 202 residues

DTPFADVEIA PDSGLTLLST GRESQSSSFS LVLSGRMRAS TGTELNGEP IKATKLAKHV ALAGIPEIDS  
LERLVTVRTV VREQLAWSSP WYLMVPRDIS DSGRWVDVEK HGLNLNPKT LIGDLSVLER FKLRIALALL  
ARPEAQLLVV DDPDQVRSME LRAEVLHALK GVAEDLPVVV VSTNPDEFDSL ADTALTITGA GN

>RXA01604-nucleotide sequence B: coding region

GACACACCCTTCGCCGATGTTGAGATAGCTCCAGACAGCGGACTCACTTTGCTGAGCACCGGGCGGAATCCCAATC  
CAGTTCCTTTTCTTTGGTACTTTCCGGCCGCATGCGCGCCTCCACCGGAACCATCGAATTAAACGGCGAACCCATCA  
AGGCAACCAAGCTGGCCAAGCATGTGGCTTTGGCGGGCATCCCTGAAATCGATTCACTCGAGCGACTTGTCACTGTG  
CGCACC GTTGTCCGTGAACAACCTCGCCTGGTCAAGCCCTTGGTACCTGATGGTGCCCGAGGGATATTAGTGATTCCGG  
ACGGTGGGTTGACGTGCAAAAGCATCTTGGCCTGAACCTGAACCCTAAACCTTAATCGGCGACCTCAGCGTGCTCG  
AGCGTTTTAAGCTGCGCATCGCGCTGGCGCTGCTGGCGCGGCCAGAGGCGCAACTGTTGGTCGTGGATGATCCCGAT  
CAAGTGCGCAGCATGGAATTGCGTGCGGAGGTGTTGCACGCATTGAAAGGCGTTGCAGAGGATCTCCCTGTGGTCGT  
GGTATCCACCAACCCAGATTTGATTCCTTGGCCGATACCGCTTTGACCATTACGGGGGCTGGAAAC

>RXA01604-nucleotide sequence C: downstream

TAATGGCATTITTTTACACTTTGGC

>>RXA00456-amino acid sequence

(1-312, translated) 104 residues

VLQALLAIMV SLSVAAILEG NRALVGLLLA TTLGLGVAQW IQKVVAEDLG QHYVHEVRRE LVGAALVPGN  
TASLGVTVTR ASNDLTAVRN WVALGIVPMV TGLP

>RXA00456-nucleotide sequence A: upstream

CTCACCAACCCGGAGATCGTCACAGCGGTGCTAACGGATCATGCCTAGCTTATGGCGTGCTCGTCGCAGACTTTTGC  
TCATTGCCCTAGGTGTACTTGGT

>RXA00456-nucleotide sequence B: coding region

GTGCTGCAGGCACTGCTGGCGATCATGGTGTCGTTGAGCGTAGCCGCCATACTTGAGGGAAACCGAGCACTTGTTGG  
ATTGCTGCTTGCTACCACGTTGGGTTTGGGGGTGGCGCAGTGGATTCAAAAAGTAGTGGCAGAAGATCTAGGCCAGC  
ATTATGTGCATGAGGTGCGTCGTGAATTGGTGGGTGCTGCGCTGGTGCCTGGAAATACGGCCTCGTTGGGCGTGACT  
GTCACCCGAGCCAGCAATGATCTCACCGCGGTGCGCAATTGGGTGGCTTTGGGCATTGTTCCGATGGTCACCGGGCT  
GCCG

>>RXA00243-amino acid sequence

(1-927, translated) 309 residues

VTSEQALDPI HPGQFRLSRI QLINWGTfHG TVDIPVTREG ILVTGGSGSG KSTLIDAITA VLLPQGKLRf  
NSAAQANTPR NKGRSLVTYI RGAWRAQEDP LQDQIVSTYL RPRATYSLVG LTYSNGEGVE HTLVAlFYLK  
SGHNLTSDIS SYYGVPFVDQ DINALLDFLK EGIDKRQIRA AFKEAlFSEQ HSVFSGRFRS RLGISSEEAL  
LLLHRAQSAK DLQSLDDLFR DYMLVEPDTF SIAKTAVEQF QDLEGAYEQV EDIKRQIHTL DPLVQLKNRR  
EKAQQSKDHA NALKKALPTV GNRIKKEEQ

>RXA00243-nucleotide sequence A: upstream

CACTGCGCCAGATTTTTGATGCCGACACTGTGGCAGGTGTGCGCGCTGAGTACGAAAAATTTAACAAAGCAGCCCAT  
GATGGAAATGAAGAGGAACAGAA

>RXA00243-nucleotide sequence B: coding region

GTGACCAGCGAACAAAGCTTTAGATCCTATCCACCCAGGTCAGTTCGGTCTTTCTCGGATTCAGTTGATCAACTGGGG  
AACCTTCCACGGAACGGTGGACATTCTGTGACCAGGGAAGGAATCTTAGTTACCGGTGGTTCGGGATCAGGAAAAT  
CCACGCTGATTGATGCGATCACGGCGGTATTGCTTCCGCAAGGAAAGCTGAGGTTTAACTCTGCCGCACAGGCTAAT  
ACTCCGCGGAATAAGGGACGCASTTTGGTTACCTATATCCGTGGCGCTTGGCGTGCGCAGGAGGATCCGCTGCAGGA  
TCAGATTGTCTCCACGTACCTACGTCCCCGCGCAACCTATTCTGCTGGTTGGATTGACTTATTCCAACGGTGAAGGCG  
TCGAGCACACCTTGGTGGCTATTTTCTATCTGAAATCGGGACACAATTTAACCTCCGATATTTCTTCATATTATGGT  
GTGTTTCCCGTTGATCAAGACATCAATGCGCTGCTGGATTTCTGAAAGAGGGCATCGATAAACGCCAGATCAGAGC  
TGCTTTCAAGGAAGCCATCTTTAGCGAGCAGCATTCTGTATTCTCCGGCAGGTTTAGAAGCCGTTTGGGGATCTCCA  
GTGAGGAAGCTTTGCTGTTGTTGCACCGCGCGCAGTCGGCGAAAGATCTTCAAAGCTTGGATGATCTATTTCGGGAT  
TACATGCTGGTGGAACCGGATACGTTTACGATTGCCAAAACCTGCCGTGGAACAATTCCAAGACCTTGAAGGTGCTTA  
TGAGCAGGTCGAAGATATTAAACGGCAGATCCACACCCTGGATCCTTTGGTGCAGCTGAAGAATCGGCGAGAGAAAG  
CGCAACAGTCCAAAGATCATGCCAATGCACTGAAGAAGGCGCTGCCGACTGTCCGGAATCGCATTAAGAAGGAAGAG  
CAA

>>RXA00259-amino acid sequence  
(1-2202, translated) 734 residues

MSGLETFPSD AAKNNTVKTD GDSVSGRDLP ITKISEDRE RSAYSAQLAN IICDVAPWGA STVFSLTGQW  
GSGKTSLVNL IRSEESLSNE KWTIVDFNFW VASDPQSLIE EFYRVIVGTV PDDKTGQKIK TVLQKTFSTI  
GSIAGGVGGF GVLEALALSK GVDAANAVYK TWKQEQDSWP TLYTRAANHF KDLNKRILIV VDDIDRLHTD  
ELALLMKVIR LLGRFPQVNY LLVYEEESLL TTLARSTAVG GSEDDALRFM EKIVQYPFDV PPLTSFQIEK  
ELSALFDKLF QGVSLSGDPE DFALVKSRMF DVWEKTLVTP RLLHRFAALL TNWTRIIYSG EVNGVDLTIL  
ATIRIVFPSV YKRLSRAKEV LLQGGRTTGS QKPGWEKQLC EGMNNEQMDL LKTMLLFLFP RLSDHPSTRM  
HRERGISTEV YFDTYLMFOR PGHVISDEQL DKYLSNADDA MGFVDLINS DNDMVASVMK KLPLAIDRLD  
GEGVRHMAVE VLFTAANGMH DKGRQVRMSG IFSDLYSHAC SILGALPOLP VEQLYEKFFS EMTLINEAAFW  
LNQVGERARA CGNDVSGLEL FRKVNKTEA RILSVLKNQD PSDWDLGPYS LGILAKSSNF SSVLKSLSQSG  
IEEHQFDVID IGVLFLLTVY SSRQGPSGGA WIDSFQHS LF SRYVPDSLRA ITKSEVDVEL GKIQTDFSW  
EGKRKV VAYA LETGRSDFTR ERLGGYSIAD SIVD

>RXA00259-nucleotide sequence A: upstream

GGCCTTATTAACATACGGCGGTTCTAGCACACACGCGATGGATGATGTGTGCCACCGATATTGCTGCATGTGCCAAGA  
TTTAGCACGGTACAGTGCTAGAA

>RXA00259-nucleotide sequence B: coding region

ATGAGCGGACTGTTTACCCCATTTTCAGATGCGGCAAAAACAACACGGTAAAACTGATGGAGATTCAGTATCTGG  
TCGAGACTTGCCTATTACTAAGATCTCTGAGGATCGTTTCGAGCGTTCTGCGTATTTCAGCCCAGCTGGCAAATATAA  
TCTGCGATGTGGCACCTTGGGGAGCGAGCACTGTTTTAGTCTTACTGGTCAGTGGGGCAGTGGTAAGACATCTCTT  
GTTAATTTGATTTCGCTCGGAAGAATCTCTATCGAACGAAAATGGACAATCGTTGATTTCAACCCGTGGGTGGCCTC  
TGACCCGCAATCTTTGATTGAGGAGTTTACCGAGTAATCGTTGGGACGGTACCTGATGATAAGACCGGCCAAAAGA  
TCAAACTGTTCTGCAGAAAACCTTTAGCACGATTGGGTCAATTGCAGGTGGGGTCGGAGGGTTTGGTGTCTTAGAA  
GCACTTGCGCTCTCAAAAGGAGTAGATGCTGCAACGCTGTATATAAGACATGGAACAGGAGCAAGATTCGTGGCC  
AACCTGTATACAGTGTCTGCGAACCATTTTAAAGATCTGAACAAGCGAATTCTCATTGTCTGCTGATGATTTGATC  
GCCTCCATACCTGATGAATTGGCGCTGTTAATGAAAGTAATACGCTTGCTTGGACGATTCCCGCAGGTGAATTATCTT  
TTGGTTTGAAGAAGAATCACTGTTAACGACGCTAGCCAGATCGACAGCTGTAGGTGGTAGCGAAGATGATGCTTT  
GCGTTTCATGGAGAAAATCGTGCAGTATCCTTTTCGATGTTCCGCTCTGACATCATTTCAAATAGGAAAAGAGCTCA  
GTGCATTATTTGACAAGCTTTTCCAGGGTGTTCGCTATCGGGTGATCCTGAAGACTTTCACTAGTAGGAGCTGAGA  
ATGTTTCGATGTCTGGGAAAAGACTCTGGTCACGCCGAGGCTGTTGCACCGTTTGTCTGCTCTACTAACCACTGGAC  
TCGGATATATGGATCAGGTGAAGTTAACGGCGTTGATCTCACAATACTTGCGACCATTCGAATTGTTTTTCCGCTCTG  
TGTATAAACGCTCTTCTCGAGCGAAGGAAGTATTGCTTCAAGGAGGTGCAACGACAGGCTCGCAGAAAACCGGTTGG  
GAAAAGCAATTATGTGAGGGGATGAACAACGAGCAGATGGATCTTTTAAAGACCATGCTTTTTGTTCCCTTTTCCACG  
TCTTTCGGATCACCTAGTACGAGAATGCATCGTGAGAGGGGGATCTCGACGGAAGTTTATTTTGACACGTACCTCA  
TGTTTCAAAGACCTGGACATGTCATAAGTGATGAACAGTTGGATAAGTATCTATCTAATGCGGACGATGCTATGGGT  
TTCGTGATTTAATTAAGTCCGATGACAATGACATGGTGGCATCAGTGATGAAAAAGCTTCTCTAGCAATTGATCG  
ACTTGATGGAGAGGGTGTAGGCACATGGCAGTTGAGGTGTTATTCACCGCTGCTAATGGTATGCATGATAAAGGTC  
GTCAAGTGCGTATGAGCGGCATATTCAGTGACCTGTATTCCCATGCGTGCTCGATTCTTGGTGCATTGCCCTCAATTA  
CCAGTGGAACAACCTCTATGAGAAATTCTTTTCTGAGATGACGCTTAATGAGGCTGCTTTCTGGTTAAACCAGGTGGG  
GGAAAGGGCTAGAGCCTGTGGTAATGATGTAAGTGGCCTTGAGCTTTTTCTGTAAGTTAATATAAAGACCGAAGCTA  
GAATTTTAAGTGTATTGAAGAATCAGGACCCCTCAGATTGGGATTTAGGTCCATATTGCTTGGTATTTTGGCGAAA  
AGCTCGAATTTTCTTCAGTGCTGAAGTCTCTGCAAAGTGGTATAGAGGAACATCAGTTTGATGTGATAGATATTGG  
AGTGCTTTTCTTAACGACTGTGTATTCTTCGCGACAGGGACCAAGCGGTGGTGCATGGATAGATTCTTTTCAGCATA  
GTCTGTTTTACGGTACGTACCTGATTCTCTACGGGCTATAACCAAGTCTGAAGTAGATGTAGAAGTAGGTAAGATA  
CAGTTCACGGATTTTAGCTGGGAAGGGAAGCGAAAAGTTGTCCGATATGCACTGGAGACTGGAAGAAGTGATTTTAC  
TCGAGAACGATTAGGGGGCTACAGTATCGCAGATTCTATAGTCGAT

>RXA00259-nucleotide sequence C: downstream

TGATGAGGCTGAGGTCATGACTT

>>RXA00164-amino acid sequence  
(1-1689, translated) 563 residues

VGRIPRAKWW FLGALVLLSA GAYASVLVPQ VLGRIVDLVS DGAQMRDFVE LSVILIAVAI AGAVLSACGF  
YVVSRISEKI IANLREDMVG TALGLPTHQV EDAGSGDLVS RSTDDVSELS AAVTETVPIL SSSLEFIAAT  
IIALFSLDWQ FVLIPVVVAP VYYFASKHYL SKAPDRYAAE RAAMAERARK VLEAIRGRAT VRAYSMEDAM  
HNQIDQASWS VVVKGIRART TMLILNMWML FAEFLMLAVA IVIGYKLVID NALTIGAVTG AVLMIIRLRG  
PMNMFMRVLD TIQSGYASLA RIVGVVADPP IPVPSGVKA PQGKVELRNV SFSYGDSWAV KDIDITINS  
ETVALVGASG AGKTTVAALL AGLRVPDQGO VLVDDFPVSH LSDRERIARL AMVSQEVHVF SGTLRQDLTL  
AKPDASDEEL AHALGQVNAL DWLESLPEGL DTVVGARGIQ LEPVVAQQLA LARVLLNPA IVIMDEATAE  
AGSAGASALE EAADAVSKNR SALVVAHRLD QASRADQILV MDKGEVVESG THQELLDHGG IYQRLWTAWS  
VGR

>RXA00164-nucleotide sequence A: upstream

CTGCTTTGCGGGAGGTTATGAAATGAGTGGGGAGACGTGCGAAAGCATGCGCTTTCCGTTGGCCAGCCTGCCGCAAG  
TGCGGCGCGAGGTGGCCCGGCAG

>RXA00164-nucleotide sequence B: coding region

GTGGGTGCTATTCCGCGGGCGAAGTGGTGGTTTTTAGGCGCGCTGGTGTGCTGAGTGCGGGCGCTTATGCGTCGGT  
GCTGGTGCGCGAGGTGCTGGGGCGGATTGTGGATCTGGTGTCCGATGGCGCGCAGATGCGTGATTTTGTGAGCTCA  
GTGTGATTCTCATTGCGGTGGCAATTGCGGGCGCGGTGCTCAGTGCGGTGCGGGTCTATGTGGTGTGCGGGATTCT  
GAGAAGATTATCGCCAATTTGAGGGAAGATATGGTGGGCACCGCGCTTGGGTTGCCACGCACAGGTGGAAGATGC  
GGGCTCTGGCGATTGTGGTGACCGCTCCACCGATGATGTCTCCGAGCTATCCGCAGCGGTGACAGAGACCGTCCCGA  
TTTTAAGTTCCTCACTGTTTACCATTGCCGCGACGATTCGCGCTGTTTTCTTTGGACTGGCAATTTGTGCTCATT  
CCTGTGCTGGTGGCGCGCGGTGTACTACTTCGCGTCCAAGCACTATTTGAGCAAGGCGCGGATCGGTATGCGGCAGA  
ACGCGCGCGCATGGCGGAGCGTGCAGCGAAAGGTACTTGGAGCTATTCGCGGGCGTGCAACTGTGCGGGCGTATTCCA  
TGGAAGATGCCATGCATAATCAGATTGATCAGGCGTCTGGTCTGTGGTGGTCAAGGGTATTGCTGCGCGCACCA  
ATGTTGATTTTGAACATGTGGATGCTGTTTGCAGAAATTCCTCATGCTCGCGGTGCGGTGGTGTGATCGGCTACAAGCT  
GGTCATTGATAATGCGCTGACGATCGGCGCGGTTACCGGTGCGGTGCTGATGATTATTGCTCTGCGTGGCCCGATGA  
ATATGTTTCATGCGCGTGCTCGACACCATTCATCCGGCTATGCGTCTGCTGGCGCGCATCGTGGGAGTTGTTGCGGAT  
CCGCCGATTCTGTGCGCGACAGCGGTGTGAAAGCACCTCAGGGCAAAGTGGAAATTGCGCAACGTGAGCTTTAGCTA  
TGCGGATTCTTGGGCGGTGAAAGACATCGACATCAGATCAATTCCGGCGAAACTGTGCGGCTCGTGGGCGCATCTG  
GCGCAGGTAAGACGACGGTGCAGCGCTTGTGCGGGGCTTGGGGGTGCCAGATCAAGGGCAAGTGCTTGTGACGAC  
TTCCCCGTCTCTACCTCTCTGACCGCGAGCGTATCGCCCGCTTGGCCATGGTCAGCCAGGAGGTTTATGTTTTCTC  
CGGCACGCTGCGCCAGGATCTCACCTTGGCTAAACCAGATGCCTCCGATGAGGAATTAGCGCATGCTCTTGGGCAAG  
TTAATGCCCTTGAAGTGGTGGAGAGTCTTCCAGAAGGACTGGACACGGTGGTGGTGGCGGAGGAATCCAGCTAGAA  
CCAGTGGTGGCTCAGCAGTTGGCGTTGGCCCGGTTGTTGTTGCTCAATCCGGCGATCGTCATCATGGATGAAGCCAC  
GGCAGAAGCAGGATCGGCGGGTGCCAGCGCACTGGAAGAGGCTGCAGATGCAGTGAGCAAGAACCCTTCCGCATTGG  
TGGTGGCGCACCGGTTGGATCAGGCATCGCGGGCTGATCAGATTCTGGTGTGATAAGGGGGAGGTTGTGGAATCC  
GGTACTCACCAGGAGTTATTGGATCACGGGGTATTTATCAGCGTCTGTGGACTGCGTGGAGTGTGCGGAAGA

>RXA00164-nucleotide sequence C: downstream

TAGTTGACTGTTCAATGCGTTGA

>>RXA00165-amino acid sequence  
(1-1416, translated) 472 residues

VASAGMAASF ICNGLTPVIV GKAVDEAIGT SDLQRLWFI AMLAVLEFLTA MTVNWIARYM LVRSQQLVSH  
DLRMLVTDRI QDPRGFAGKE RTAGGLLSIA SSQTQRVGDI VMTVFPPVAE LASIIYGAVV MYSINPWLSV  
AVLIGGPLLV VVAIQVSKPL QKRSGARQQA VAQAAATATD VVQGLRILKG LGAIVTVRRR YEASIGEAYR  
KTVHADAAEA RLNGVTDAAG AIFVSALGIG AGFLALQGM SIGDLITVVG LTQFLIMPMT MLGRNVASRW  
ASAEASAKRI RGVLGADFER VSAHDADKAE EIIQQLAKGL TVIRGTDEQL VEVLEQLPRT RVIVAPHAAD  
LFDQSVRDNV HPVAEVAEKA IEVASCDDIP GGSSKIVGEG GRLLSGGQORQ RVALARIAAF DPEVLVLQDP  
TTAVDSVTEQ NIAQOVAHR AGKVTIVFSE APAWSAVADQ HVEAAALREV MK

>RXA00165-nucleotide sequence A: upstream

AAACCTCCCCGGGCGCGCGCGACCGTCCAAGATGCCGGCGTTGGATGCCAAATTATGGACTCTCAAAGTGGCGT  
TGTCGCAGCGGCCGTGGAGCTTT

>RXA00165-nucleotide sequence B: coding region

GTGGCGTCTGCTGGCATGGCGGCGTCTTTTATCTGCAATGGGTAAACGCCTGTGATTGTGGGTAAGGCGGTGGATGA  
GGCTATTGGCAGCAGCGATCTGCAGCGATTGTGGTCTGGATTGCCATGTTGGCGGTTCTTTCTTAACGGCGATGA  
CGGTGAACCTGGATTGCTCGGTACATGTTGGTGGGAGCCAGCAGTTGGTCAGCCATGATTGCGCATGTTGGTGACT  
GATCGGATTCAAGATCCGCGTGGTTTTGCTGGAAAAGAGCGCACTGCCGGTGGATTGTTGTCGATTGCGTCATCGGA  
TACGCAGCGGGTGGGCGATATCGTCATGACGGTGTTCCCGGTGGCGGAATTGGCGTCGATTATTTATGGCGCCG  
TGGTGATGTACAGCATTAAATCCGTGGTTGAGTGTGGCTGTGCTGATTGGTGGACCGCTGCTGGTTGTGGTGGCTATT  
CAGGTCTCAAAGCCGTTGCAGAAGCGTTCGGGTGCTCGTCAGCAGCGGTGGCACAGGCTGCGGCTACTGCAACTGA  
TGTGGTGCAGGGCTTGAGAATTTTGAAGGGTTTGGGCGCGATTGTACGGTGCGCCGTCGGTACGAGGCGATTCTG  
GTGAGGCTTATCGGAAGACGGTTCATGCGGATGCTGCGGAAGCTCGCTTGAATGGTGTACCGATGCGGCGGGCGCC  
ATCTTTGTGTCCGCGTTGGGTATTGGAGCAGGATTTTGGCGCTGCAAGGTCAGATGAGTATTGGTGATTTGATCAC  
GGTTGTGGGACTCACACAGTTTTTGTATCATGCCGATGACCATGCTTGGTTCGAAATGTGGCATCGCGCTGGGCATCGG  
CGGAGGCGTCGGCAAAGCGTATTAGGGGAGTGCTCGGTGCTGATTTTGTAGAGAGTGTCTGCGCATGATGCGGACAAG  
GCTGAGGAGATTATCCAACAACCTGCCAAAGGTTTGACGGTTATTTCGAGGCACTGATGAGCAGCTCGTTGAGGTATT  
AGAGCAGTTGCCACGTACTCGGGTGATTGTGGCTCCTCATGCGGCGGATCTTTTTGATCAAAGTGTGAGGACAATG  
TGCATCCCGTGGCAGAGGTGCGGAGAAAAGCCATTGAAGTTGCCTCCTGTGACGATATTCCAGGTGGTAGTTCCAAG  
ATTGTGGGCGAGGGTGGACGGTTGCTCTCGGGTGGTCAGCGTCAGCGCGTTGCACTGGCTCGGGCGATTGCTTTTGA  
TCCAGAGGTGTTGGTGCTTCAAGATCCCAACGGCAGTGATTCTGTGACGGAGCAAAACATTGCTCAGCAAGTGG  
CAGCACACCGTGCAGGAAAAGTGACCATTTGTGTTTAGTGAGGCACCCGCGTGGAGTGGGTGGCTGATCAACACGTT  
GAGGCAGCTGCTTTGCGGGAGGTTATGAAA

>RXA00165-nucleotide sequence C: downstream

TGAGTGGGAGACGTCGAAAAGC

>>RXA00188-amino acid sequence  
(1-1443, translated) 481 residues

LGEFIYTMKN VRKAIGDKVI LDNVYMSFYP GAKIGVVGP N GAGKSSILKI MAGLDQPSNG EAFLDPGATV  
GILLQEPPLN EEKTVRQNV E GLGEIFEKK QRFDAIAEEM ATNYTDELME EMGVLQEALD AADAWEIDSK  
IDQAMDALRC PPSDEPVTHL SGGERRRVAL AKLLLAEPDL LLLDEPTNHL DAESVLWLEK HLADYKGAVL  
AVTHDRYFLD HVAQWICEVD RGQLHPYEGN YSTYLEKKAE RLEVSGKKDQ KLQKRLKDEL AWVRSGAKAR  
QAKNKARLQR YDEMVAEAEK YRKLDFFEEIQ IPTPPRLGNK VVEVKDLEKG FDGRVLIKDL SFTLPRNGIV  
GVIGPNGVGK STLFKTIVGL ENPDAGSVEI GDTVQLSYVD QGRENIDPEK TVWETVSDGL DYIIIVGQNE M  
PSRAYLSAFG FKGADQQKPS KVLSGGERNR LNLALT LKQG GNLILLDEPT NDLDETGLS L

>RXA00188-nucleotide sequence A: upstream

TGCTGTAAGGCCCACTTGCCCGAGGTCCCTAAATTGGCGCTGCTCATGTAATCTTGTGCGTTGAGTACTTAAACGCA  
AATTTCGATTTAAAGGGGTACA

>RXA00188-nucleotide sequence B: coding region

TTGGGCGAGTTCACTACACGATGAAGAAGCTGCGTAAGGCAATCGGTGACAAGGTCATCTTGGACAATGTCTACAT  
GTCCTTCTACCCAGGCGCCAAGATCGGTGTCGTCGACCCAACGGTGCCGGTAAGTCATCGATCTTGAAGATCATGG  
CTGGGTTGGATCAGCCTTCCAACGGTGAAGCATTCTTGGATCCAGGTGCAACCGTGGGCATCCTGTTGCAGGAGCCA  
CCACTAAACGAAGAAAAGACTGTTTCGCCAGAACGTCGAAGAAGGCTTGGTGAGATCTTCGAAAAGAAGCAGCGCTT  
CGACGCCATCGCAGAAGAAATGGCAACCAACTACACCGACGAGCTCATGGAAGAAATGGGCGTTCTGCAGGAGGCTC  
TCGACGCAGCTGATGCGTGGGAGATCGACTCCAAGATCGACCAAGCAATGGATGCATTGCGTTGCCCTCCAAGCGAT  
GAGCCAGTTACCCACCTCTCCGGTGGTGAGCGTCGCCGCGTCGCACTGGCAAAGCTTCTGCTTGACAGCCAGATCT  
GCTGCTGCTCGATGAGCCTACTAACCACCTTGACGCCGAGTCCGTCCTGTGGCTGGAGAAGCACCTCGCTGACTACA  
AGGGCGCTGTCCTTGCCGTCACACAGACCGTTACTTCCTCGACCAAGCTTGACAGTGGATCTGTGAAGTTGACCGT  
GGTCAGCTGCACCCTTACGAAGGCAACTACTCCACCTACCTGGAGAAGAAGGCAGAGCGCTCGAAGTTTCTGGCAA  
GAAGGATCAGAAGCTGCAGAAGCGCCTGAAGGACGAACCTCGCATGGGTTTCGTTCCGGCGCTAAGGCACGTGAGGCTA  
AGAACAAGGCTCGTCTGCAGCGTTACGACGAGATGGTTGCGGAAGCTGAGAAGTACCGCAAGCTCGACTTCAAGAA  
ATTCAGATCCCAACCCCAACGCGCTGGGCAACAAGGTCGTGGAGGTCAAGGACCTGGAGAAGGGCTTCGATGGCCG  
CGTGCTGATCAAGGACCTGTCTTCACCTTGCTCGTAACGGCATCGTCGGCGTGATCGGACCTAACGGTGTGGGTA  
AGTCCACCCTGTTCAAGACCATCGTTGGCCTTGAAAACCCAGACGACGAGCGTGGAAATCGGCGACACCGTCCAG  
TTGTCCTACGTGGACCAAGGCGGTGAAAACATTGATCCAGAAAAGACCGTTTGGGAAACCGTTTCCGACGGACTCGA  
CTACATCATTGTTCGACAGAACGAAATGCCATCCCGCGCATACCTGTCCGCTTCGGATTCAAGGGCGCAGATCAGC  
AGAAGCCATCCAAGGTCCTCTCCGGTGGTGAGCGCAACCGCCTCAACCTCGCGCTGACCTTGAAGCAGGGCGGAAAC  
CTGATCCTCCTCGATGAGCCTACAAACGACCTCGACGTGGAAACTCTGGGCTCCCTG

00 07 99

### Claims

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding an MCT protein, or a portion thereof.
2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule encodes an MCT protein involved in the production of a fine chemical.
3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.
4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B.
5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B.
6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.
7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A.
8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.
9. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-8 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.
10. A vector comprising the nucleic acid molecule of any one of claims 1-9.
11. The vector of claim 10, which is an expression vector.
12. A host cell transfected with the expression vector of claim 11.
13. The host cell of claim 12, wherein said cell is a microorganism.
14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.
15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.



- 5 16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.
17. A method of producing a polypeptide comprising culturing the host cell of claim 12 in an appropriate culture medium to, thereby, produce the polypeptide.
- 10 18. An isolated MCT polypeptide from *Corynebacterium glutamicum*, or a portion thereof.
- 15 19. The polypeptide of claim 18, wherein said polypeptide is involved in the production of a fine chemical production.
- 20 20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.
21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof.
- 25 22. The isolated polypeptide of any of claims 18-21, further comprising heterologous amino acid sequences.
- 30 23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A.
- 35 24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.
25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.
- 40 26. The method of claim 25, wherein said method further comprises the step of recovering the fine chemical from said culture.
- 45 27. The method of claim 25, wherein said method further comprises the step of transfecting said cell with the vector of claim 11 to result in a cell containing said vector.
28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.
29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum*, *Corynebacterium herculis*, *Corynebacterium lilium*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*.

09.07.99

5 *Corynebacterium acetophilum*, *Corynebacterium ammoniagenes*, *Corynebacterium fujiokense*, *Corynebacterium nitrilophilus*, *Brevibacterium ammoniagenes*, *Brevibacterium butanicum*, *Brevibacterium divaricatum*, *Brevibacterium flavum*, *Brevibacterium healii*, *Brevibacterium ketoglutamicum*, *Brevibacterium ketosoreductum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium paraffinolyticum*, and those strains set forth in Table 3.

- 10 30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.
- 15 31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.
- 20 32. The method of claim 25, wherein said fine chemical is an amino acid.
- 25 33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.
34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.

CORYNEBACTERIUM GLUTAMICUM-GENE, DIE PROTEINE CODIEREN, DIE AN DER MEMBRANSYNTHESE UND AM MEMBRANTRANSPORT BETEILIGT SIND

## 5 Zusammenfassung der Offenbarung

Isolierte Nukleinsäuremoleküle, die als MCT-Nukleinsäuremoleküle bezeichnet werden und neue MCT-Proteine aus *Corynebacterium glutamicum* codieren, werden beschrieben. Die Erfindung stellt zudem

- 10 Antisense-Nukleinsäuremoleküle, rekombinante Expressionsvektoren, die MCT-Nukleinsäuremoleküle enthalten und Wirtszellen, in die die Expressionsvektoren eingebracht worden sind, bereit. Sie stellt weiterhin isolierte MCT-Proteine, mutierte MCT-Proteine, Fusionsproteine, antigene Peptide und Verfahren zur Verbesserung
- 15 der Produktion einer gewünschten Verbindung aus *C. glutamicum* bereit, die auf der genetischen Manipulation von MCT-Genen in diesem Organismus beruhen.

20

25

30

35

40

45

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**